

Reaching for the Moon

Naive or not, NASA's next shot at landing on the Moon can succeed only if it is launched as a genuinely international collaboration.

In 2018, according to a timetable announced last month, the United States will send astronauts to the Moon for the first time since 1972. Four people would stay on the lunar surface for up to a week, having arrived in a new lander attached to a new crew transport launched by rockets derived from the space shuttle. Eventually they would live for six months at a time in Antarctica-style outposts.

The estimated price tag to develop all this new hardware is \$104 billion between now and the first landing. Or rather, the seventh lunar landing as NASA likes to call it, to emphasize continuity with the past. The goal this time is not just flags and footprints, not just beating the Russians, but the beginning of humanity's permanent expansion into the Solar System. To even talk in such terms implies a particular view of human progress that some find inspiring and others dismiss as almost childish. In a time of war, hurricanes and soaring energy prices, is shooting for the Moon optimism or hubris? Either way, NASA seems to be set on this particular course.

Given that public opinion is divided on the subject, and that there's no real rush to return to the Moon, the space agency has a responsibility to execute the idea with as little waste as possible. That will require a major change of tack at NASA, as well as bold new approaches to both domestic and international politics.

On the domestic front, Congress needs to back off from the parochial meddling that has long contributed to NASA's inefficiency. Senators from Texas and Florida, where key NASA centres are located, are already trying to fend off cuts to the space shuttle and space station that are needed to pay for the Moon missions. NASA administrator Mike Griffin, who has an engineer's instinct for efficiency, has said that NASA's workforce will remain about the same size as it is today. But the agency may need skills in new areas, and the jobs may be in different congressional districts. Griffin needs the freedom to make these decisions based on his practical needs, not on political considerations.

Nor should the United States try to go it alone to the Moon. Japan and India are taking their own first (robotic) steps in the same

direction in 2007, with scientific missions sent to lunar orbit. So is China, which is also building up a modest capability to launch people into space. Europe and Russia are making their own cautious plans for lunar exploration as they watch NASA's plans unfold.

All these partners are interested in building up their own domestic capabilities in space, so a certain amount of duplication of effort is inevitable. But to every extent possible, the construction of a lunar base should be an international venture that takes advantage of each partner's strengths and interests. Canada and Japan might emphasize robotics, for example. Russia builds reliable spacecraft and rockets. A lunar programme should include no more overlap than is required to ensure a back-up for essential technologies.

The International Space Station has hardly been an inspiring model for such an enterprise. So far the coalition that is building it has held together — but NASA's partners in Europe, Japan and Canada are still nervous over whether the United States will renege on its commitment to launch their modules. It is unclear, to put it mildly, that any of the partners will get their money's worth.

Building a permanent human presence away from Earth is a far more daunting venture, and can't be handled in the same way, with a single memorandum of understanding between the international partners. The collaboration will take place in many shapes and forms over decades, and must, therefore, be truly collaborative in both spirit and practice, in a way that the US-led space station has never been.

If we are to accept the high-minded premise that humanity is poised to take its next evolutionary step, then the politics of the Moon programme should be high-minded too. The alternative is to admit that this is just another pork-barrel project. The onus remains on Griffin, the Bush administration and their prospective international partners to show that it will be any more than that. ■

"The construction of a lunar base should be an international venture that takes advantage of each partner's strengths."

In need of rehab

The reputation of one of the world's most respected regulatory agencies is on the wane.

The US Food and Drug Administration (FDA) is in trouble. Last month's abrupt resignation of its commissioner, Lester Crawford, leaves the agency again bereft of leadership as it struggles to absorb the aftermath of last year's traumatic withdrawal of the painkiller Vioxx.

The Vioxx episode has left the agency in crisis, facing immense,

conflicting pressures from watchdog groups and the pharmaceutical industry on how it should handle drug approvals. In these circumstances, the agency needs a commissioner who can rise above the political fray and convince the public that the FDA is in safe hands, while taking a sophisticated and innovative approach to drug approval. Unfortunately, there is little sign that it's going to get one.

The reasons for Crawford's departure, only two months after his confirmation in the position by the Senate, remain murky. The timing of the announcement — on a Friday afternoon as Hurricane Rita bore down on the Gulf coast — bore all the hallmarks of an effort by the Bush administration to bury the event (see *Nature* 437, 606; 2005).

To the surprise of the agency's supporters and detractors alike, the



Bush administration announced that Crawford would be replaced on an acting basis by Andrew von Eschenbach — who would also continue to serve as director of the National Cancer Institute (NCI), a massive research agency with a stake in some of the FDA's toughest regulatory decisions. Last Friday, von Eschenbach admitted that this would be impossible, and said he would temporarily shelve his daily duties at the NCI and excuse himself from some cancer-related activities at the FDA (see page 802).

Crawford's record as acting and then permanent head of the FDA was underwhelming. He had little public visibility and seemed reluctant to back up his own scientific advisers when their advice ran counter to Bush administration doctrine, for example to make Plan B, the morning-after contraceptive, available over-the-counter from pharmacists (see *Nature* 437, 179; 2005). Now his departure is being left wholly unexplained, prompting reports of financial conflicts, as well as a bipartisan congressional investigation.

It is not clear that von Eschenbach can do much better. The 63-year-old urology surgeon has exasperated NCI researchers by making it a goal to end suffering and death from cancer by 2015 — an improbable aim he describes as “within our grasp”. His public commitment to more rapid approval of experimental cancer treatments also deserves close examination in the light of several drug withdrawals, for safety reasons, in the past year alone.

In his three years at the NCI, he has become known as a hands-off manager who leaves the workings of bureaucracies under him largely

to subordinates. This is not the prescription for an agency that has been rocked by serious crises and that now needs a leader with a firm grasp of policy details that ultimately affect millions of lives.

Yet if past inattention is any indication, it seems likely that the White House will leave von Eschenbach — a Bush family friend — holding the fort at the FDA while it is preoccupied elsewhere. That would leave an agency that has lacked a permanent head for most of Bush's presidency in limbo yet again. Three years ago, a government survey of 400 FDA scientists found 18% of them reporting that they “have been pressured to approve or recommend approval” for a drug “despite reservations about the safety, efficacy or quality of the drug”. In the absence of firm leadership, scientists at the FDA's headquarters in the Washington DC suburbs will be left to do battle with ideologues such as Scott Gottlieb, a 33-year-old physician and former Wall Street tipster who was appointed in July as deputy commissioner for medical and scientific affairs at the agency.

Perhaps the most worrying prospect is that of an agency left to drift and further neglect under a stop-gap commissioner for another three years, until Bush has served out his term. His administration needs to find a well-qualified, permanent FDA commissioner — and soon. ■

“This is not the prescription for an agency that has been rocked by serious crises and now needs a leader with a firm grasp of policy details.”

Welcome *Nature Physics*

The launch of a new *Nature* journal comes at an exciting time for physics.

People have stopped talking about ‘physics envy’. Time was when other sciences were jealous of the prestige and funds attracted by physics, and also of its success in capturing the imagination, as it uncovered revolutionary ways of thinking about, and predicting, the constituents and governing principles of the Universe.

Nowadays, thanks to the allure of biology's progress and benefits, physics is just another discipline. But its decline in prominence should not mislead. The next generation of particle accelerators promises insights as deep as any of their predecessors, in particular in understanding the origins of mass and the symmetries underlying the laws of nature. The enduring conjugal relationship between physics and mathematics continues to stimulate both. Understanding the behaviour of electrons and light within condensed matter continues to yield not only surprises in understanding but also new technologies. And physicists' habit of thinking about the underlying questions leads them still to speculate beyond the current limits of experiment. Where does quantum mechanics fail? Is information a more fundamental quantity than hitherto realized?

It is with the enduring enticement of these challenges in mind that we welcome the launch this month of our sister publication *Nature Physics* (www.nature.com/naturephysics). It is also an indicator of success. After the Second World War, *Nature* ceased to be a vehicle for the physics community. It was only after the advent of high-

temperature superconductivity that physicists began to rediscover the journal's value. Over the succeeding two decades or so, *Nature* has re-established itself as a prime physics outlet.

At the same time, the publishing habits of physicists have also evolved. Preprint servers are now commonplace for some branches of the subject, without damaging journals. The number of papers published has grown by 3% per year, but there have been significant shifts in regional output. Between 1981 and 2001, US research output in physics fell by 1.5% (to 19,500 papers per year), Western Europe saw research output grow by 56% (to 29,100 papers), and output in Asia grew by 120% (to 22,500 papers). Within Asia, China saw its output grow from 500 articles to 5,500, Japan's grew by 67% to 11,000 and India saw a 40% increase to 2,100 papers.

Perhaps the most significant shifts are in the distribution of the physics community over that period, with the number of PhDs in physics declining markedly in the United States and Europe but increasing dramatically in Asia. *Nature* and its related journals have always had internationalism as a key ingredient, and have reflected regional growths in strength.

Experience has shown that launching sibling research journals strengthens, rather than weakens, *Nature*. More importantly, it stimulates the discipline by providing greater exposure thanks to our media and web strengths, and, above all, by providing healthy competition to established journals, to the benefit of authors and readers everywhere. *Nature Physics* is set to follow this tradition. ■

“Experience has shown that launching sibling research journals strengthens *Nature*. More importantly, it stimulates the discipline.”

RESEARCH HIGHLIGHTS

Tundra and lightening

J. Geophys. Res. **110**, G01004 (2005)

Global warming will cause parts of the Arctic to absorb up to three-quarters more sunlight than they do at present, say US researchers.

Climate scientists expect rising temperatures to cause shrubs to grow in Arctic areas that are covered in snow for much of the year. This could increase warming, because darker vegetation absorbs more solar radiation than snow. So researchers from the US Army Cold Regions Research and Engineering Laboratory and Colorado State University measured the radiation reflected from different Arctic surfaces at five sites in Alaska — the Niukluk River is shown here. They estimate that a tundra-to-shrub transition would cause an increase in absorption of 69 to 75%, depending on the latitude and time of year.



S. DOUGLAS

NEUROSCIENCE

Conscious connections

Science **309**, 2228–2232 (2005)

Consciousness fades when we fall asleep, but the brain remains active. Giulio Tononi and his colleagues at the University of Wisconsin, Madison, go some way to resolving this paradox. Using equipment that allows the human brain to be stimulated and monitored at the same time, they show that during deep non-rapid-eye-movement sleep, which occurs at the beginning of the night, neuronal firing in the brain is unable to spread across the cortex. In contrast, effective connectivity persists during quiet wakefulness; activity in one cortical area is transmitted to other areas of the brain.

DRUG DISCOVERY

Cravatt makes twin ties

Nature Biotechnol. doi:10.1038/nbt1149 (2005)

Drug developers screen libraries of small molecules to find useful biological effects, but it is not easy to identify these molecules' targets. Small molecules often bind their protein targets only weakly, making it hard to isolate the complexes from living cells.

Benjamin Cravatt of the Scripps Research Institute in La Jolla, California, and his co-workers get round this by adding a probe containing two reactive groups to the molecules that are to be screened. One reactive group attaches to the protein when the molecule binds, holding the two together. The other reacts with a fluorescent label, allowing the trapped protein to be identified.

The researchers used their method to find

a compound that inhibits proliferation in breast-cancer cells and to identify its enzyme target.

ANIMAL BEHAVIOUR

Smells fishy

Nature Chem. Biol. doi:10.1038/nchembio739 (2005)

Researchers working on strategies to control the invasive sea lamprey, *Petromyzon marinus* (pictured), have characterized the chemicals that act as a migratory cue for the creatures.

The larvae give off a pheromone that adults detect at concentrations as low as a single milligram in five Olympic-size swimming pools of water. The chemical draws the adults to suitable spawning sites upstream. Peter Sorensen, Thomas Hoye and colleagues of the University of Minnesota found three components — the most active of which is similar to the antibiotic compound squalamine, produced by sharks.

The team hopes that a fake pheromone could act as bait to help rid the North American Great Lakes of these parasitic fish, which arrived from the Atlantic a century ago.



©2005 Nature Publishing Group

MOLECULAR EVOLUTION

Sense and sensitivity

Proc. Natl Acad. Sci. USA **102**, 14338–14343 (2005)

Not all proteins evolve at the same rate, and those that evolve most slowly are often also those that are most highly expressed. Why?

Allan Drummond of the California Institute of Technology in Pasadena and his co-workers studied recent gene-chip data from yeast. They verified that expression rate is a good predictor of evolution rate for yeast proteins, then used the genome-wide data to compare explanations for rate variation. They conclude that evolution rate varies because highly expressed proteins incur translation errors more often — so their folding behaviour needs to be relatively insensitive. Resilient sequences are rare, leaving little opportunity for evolution to tinker with them.

PHYSICS

In phase

Phys. Rev. Lett. **95**, 127205; 127206; 127207 (2005)

As if superfluids weren't strange enough, quantum theory also predicts the existence of supersolids — crystalline structures that, like superfluids, will flow with zero viscosity.

Last year, researchers thought they might have detected the phase in helium-4 (*Nature* **427**, 225–227; 2004). Now three papers, from groups in Europe, India and California, describe a system that could create supersolids for study. They calculate that pouring certain kinds of superfluid into an optical lattice that holds the atoms in a triangular pattern will force them into a supersolid phase when the lattice is partially filled.

D. HANSEN

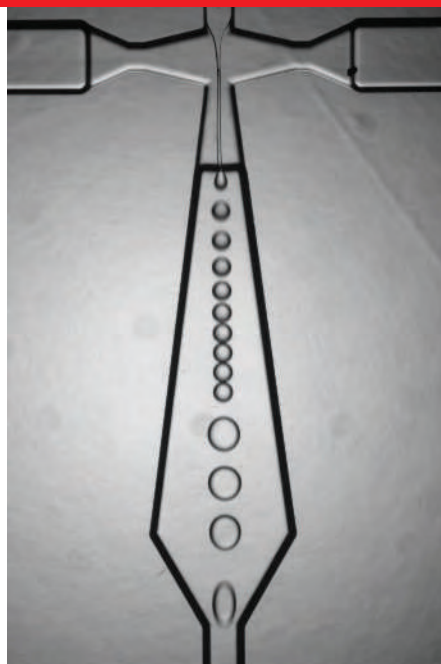
NANOTECHNOLOGY

Crystal balls

J. Am. Chem. Soc. doi:10.1021/ja051381p (2005)

A steady stream of nanodroplets is pictured (right) flowing through a microreactor designed by a team led by Richard Mathies of the University of California, Berkeley.

Unlike other systems, this set-up works even at high temperatures, and was used to create cadmium selenide nanocrystals at temperatures between 240 and 300 °C. Precursors containing cadmium and selenium are introduced into bubbles of octadecene squirted into an inert carrier fluid. The size of the bubbles, each just a few billionths of a litre in volume, is controlled by varying the injection rate of the liquids into the reactor.



ACS

BACTERIOLOGY

Hard to stomach

Lancet 366, 1079–1084 (2005)

A virulent strain of the *Clostridium difficile* bacterium that has emerged in hospitals in Canada has been characterized by Michel Warny of Acambis in Cambridge, Massachusetts, and his colleagues.

Patients who have taken antibiotics to cure another illness sometimes suffer as resistant strains of the bacterium flourish in their gut. Between 1991 and 2003, the number of people killed by a *C. difficile* infection within a month of its diagnosis increased from 4% to 13% in a region of Quebec. Warny's group identify the epidemic strain as NAP1/027, and link its virulence to elevated production of toxins.

ASTROPHYSICS

Peering into the past

Preprint astro-ph/0509303 at <http://arxiv.org> (2005)

The satellite telescope Swift is looking out for the long gamma-ray bursts produced by the violent deaths of stars. A good proportion of these bursts should come from very early stars, say Volker Bromm of the University of Texas, Austin, and Abraham Loeb of Harvard University in Cambridge, Massachusetts.

Using star formation rates, the researchers calculate that about 10% of the bursts seen by Swift will come from stars that exploded during the first billion years of the Universe's existence.

They add that some of them may even come from the first ever stars. Little is known about these bodies, which are thought to have been massive and metal-free. But to release a burst of gamma rays, they would have to have existed as binary systems in which one star stripped away the outer layer of its dying mate.

CHEMISTRY

Attack on anthrax

Angew. Chem. Int. Edn 44, 2–5 (2005)

It was recently discovered that *Bacillus anthracis* spores display a unique carbohydrate on their surface. Because vaccines made from synthetic carbohydrates have shown some potential in treating other diseases — including cancer and malaria — this tetrasaccharide has piqued interest as a candidate for a small-molecule anthrax vaccine.

Daniel Werz and Peter Seeberger, both of the Swiss Federal Institute of Technology in Zurich, report the first chemical synthesis of this tetrasaccharide. They designed the synthetic scheme so that it could be easily tweaked to modify the carbohydrate's structure — meaning that, if necessary, it should be easy to synthesize and test analogues of the natural compound.

SYNTHETIC BIOLOGY

Flak jackets

J. Am. Chem. Soc. 127, 13213–13219 (2005)

No one could expect building a living cell from scratch to be easy. But one difficult hurdle has been overcome by Jack Szostak and his co-workers at the Harvard Medical School in Boston, Massachusetts.

The team is aiming to make 'protocells' by filling vesicles made from fatty acids with ribozymes, which combine genetic and enzyme-like functions.

The problem has been that the metal ions that ribozymes typically need to function cause the fatty acids to precipitate. Now the team has found that a mixture of myristoleic acid and its glycerol ester will make vesicles that are not only immune to metal-induced precipitation, but are also permeable to magnesium ions.

JOURNAL CLUB

Günter Theißen

University of Jena, Germany

Dog domestication inspires an evolutionary biologist to reflect on the evolution of complex life.

There are some who believe that living organisms are so complex that they must have been created by an 'intelligent designer' — arguably just another term for God. Meanwhile, others argue that evolutionary biology can already fully explain how complex organisms emerged.

In my view, clarifying how complex organisms evolve remains one of the greatest challenges of biology. Rather than intelligent design, however, we need more intelligent research, such as that presented by John Fondon and Harold Garner (J. W. Fondon & H. R. Garner *Proc. Natl Acad. Sci. USA* 101, 18058–18063; 2004).

We have learned in recent years that changes in the genes that control the development of organisms can bring about major evolutionary transitions. But it has not been clear how the usually slow accumulation of random mutations could account for the fast and coordinated morphological changes seen in the fossil record.

Fondon and Gardner's study suggests a striking solution. Their research links the differing shapes of dog breeds to variation in the number of repeats of short DNA sequences in certain developmental control genes. Changes to repeat length occur much more often than other kinds of mutation, explaining how evolution can sometimes be very rapid. Also, the identified repeats encode amino acids that, ultimately, modulate the activity of genes involved in limb and skull development, explaining the effects on morphology.

I am optimistic that the authors have uncovered a crucial mechanistic detail of how genes link development to evolution. I eagerly await further analyses telling us whether this trick in dogs also works in other organisms.

SPECIAL REPORT

The 1918 flu virus is resurrected

The recreation of one of the deadliest diseases known could help us to prevent another pandemic. Or it might trigger one, say critics. **Andreas von Bubnoff** investigates whether the benefits outweigh the risks.

It is thought to have killed 50 million people, and yet scientists have brought it back to life. In this issue of *Nature*, scientists publish an analysis of the full genome sequence of the 1918 human influenza virus. And in this week's *Science*, researchers describe how they used that sequence to recreate the virus and study its effects in mice.

Some scientists have already hailed the work as giving unprecedented insight into the virus. Working out how it arose and why it was so deadly could help experts to spot the next pandemic strain and to design appropriate drugs and vaccines in time, they say.

But others have raised concerns that the dangers of resurrecting the virus are just too great. One biosecurity expert told *Nature* that the risk that the recreated strain might escape is so high, it is almost a certainty. And the publication of the full genome sequence gives any rogue nation or bioterrorist group all the information they need to make their own version of the virus.

Jeffery Taubenberger of the Armed Forces Institute of Pathology in Rockville, Maryland, is the lead author of the sequencing study. He says the work was necessary and the risks were low. The paper on page 889 gives details of the

final three genes; the sequences of the rest have already been published.

The full sequence is strong evidence that the 1918 flu virus is derived wholly from an ancestor that originally infected birds. In contrast, the viruses that caused the flu pandemics of 1957 and 1968 arose when human and avian flu viruses infected the same person at the same time, allowing their genes to mix.

All eight of the genome segments from the 1918 virus differ in important ways from other human flu sequences, suggesting that none of the genome came from a strain that had previously infected people. "It is the most bird-like of all mammalian flu viruses," says Taubenberger.

Pinpointing exactly which genetic mutations allowed the virus to jump to humans will enable scientists to recognize other bird viruses that could trigger a pandemic. Taubenberger's team has already identified 25 changes in the protein sequences of the 1918 strain that have been present in subsequent human flu viruses. These mutations are likely to be particularly important, he says. One such change, in the polymerase gene *PB2*, was found in the virus isolated from the only human fatality in a 2003 outbreak of H7N7 bird flu in the Netherlands.

In the paper in *Science* (T. M. Tumpey *et al.* **310**, 77–80; 2005), Terrence Tumpey at the Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia, and his co-workers have used Taubenberger's sequence to recreate the complete 1918 virus (see graphic).

When they used the strain to infect mice they found it was extremely virulent, and after 4 days had generated 39,000 times more virus particles in the animals' lungs than a modern flu strain (see 'How virulent is 1918 flu?'). "I didn't expect it to be as lethal as it was," says Tumpey.

The researchers compared the complete 1918 virus with strains in which some genes had been replaced by those of contemporary

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

'Fresh air' cures were used to fight flu in 1918, but reconstructing the virus may lead to more effective treatments.

strains. They found that replacing the haemagglutinin gene, which helps the virus to enter cells, made it unable to kill mice. Replacing all three of the polymerase genes, which allow the virus to replicate, significantly reduced its virulence. The haemagglutinin gene is essential, says Tumpey. "But no single change or gene is the answer," adds Taubenberger. "It's a combination effect."

Future research will involve testing reconstructed viruses with and without certain mutations, to see which are the most important for virulence. Information from this type of study will hopefully be of use in vaccine and drug design, but so far the work is more about obtaining a basic understanding of the virus than any immediate health benefits.

The studies have been praised as groundbreaking. "It's a landmark," says Eddie Holmes, a virologist at Pennsylvania State University in University Park. "Not only is this the first time this has been done for any ancient pathogen, but it deals with the agent of the most important disease pandemic in human history."

The team got permission to do the work

HOW VIRULENT IS 1918 FLU?

50 times as many virus particles are released from human lung cells a day after infection with the 1918 virus as are released after exposure to a contemporary strain called the Texas virus.

13% of body weight is lost by mice 2 days after infection with 1918 flu; weight loss is only transient in mice infected with the Texas strain.

39,000 times more virus particles are found in mouse lung tissue 4 days after infection with 1918 flu than are found with the Texas virus.

All mice died within 6 days of infection with 1918 flu; none died from the Texas strain.

FISH PHEROMONES

MADE IN THE LAB

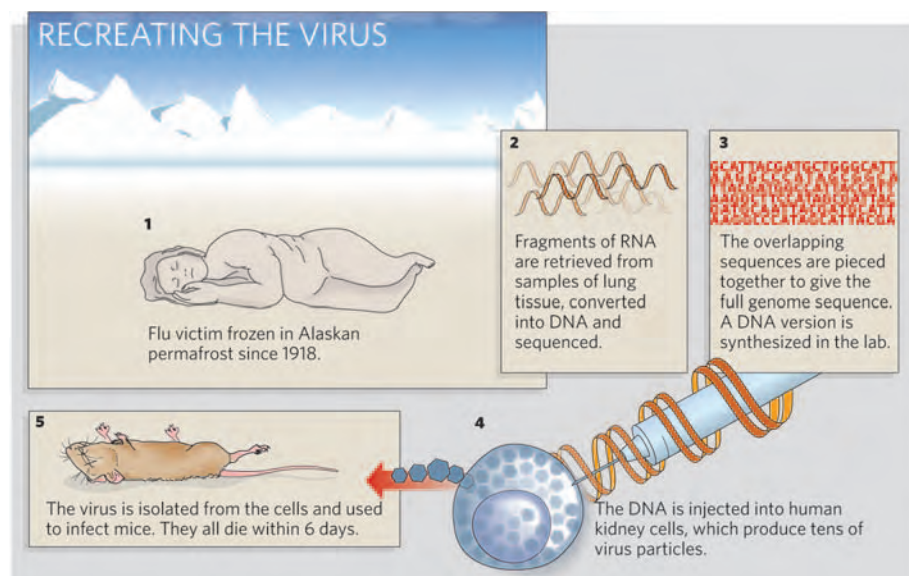
Lampreys could be lured away from Great Lakes by artificial chemicals.

www.nature.com/news

R. & T. CHURCH/SPL



IMAGE UNAVAILABLE FOR COPYRIGHT REASONS



any of these scenarios could result in a large number of deaths.

Ebright also believes that using an enhanced biosafety level-3 lab for the work was inadequate. If the researchers were going to do the work at all, they should have used level 4, the strictest biosafety condition, he says. This requires experimenters to wear full body suits. In 2003, he points out, a SARS virus escaped accidentally from a level-3 lab in Singapore, and in 2004 two further escapes occurred from such labs in Beijing.

Tumpey counters that enhanced level 3, which requires upper body suits and respirators, is safe enough. Disgruntled employees aren't a concern either, he says, because he is the only one who works with the virus. The few researchers with access to the lab undergo extensive background checks, and retina and fingerprint scans are used to prevent any unauthorized entry to the lab.

He adds that even if the virus did escape, it wouldn't have the same consequences as the 1918 pandemic. Most people now have some immunity to the 1918 virus because subsequent human flu viruses are in part derived from it. And, in mice, regular flu vaccines and drugs are at least partly effective against an infection with reconstructed viruses that contain some of the genes from 1918 flu.

Publish and be damned?

The other potential threat comes from the availability of the full genome sequence, which has been put on the GenBank database — a condition of the paper's publication. Anyone can order DNA to be made to a certain sequence, points out Jonathan Tucker, a policy analyst at the Center for Nonproliferation Studies in Washington DC. There are currently

no governmental controls on what sequences can be used, says Tucker, although some DNA synthesis companies now screen their orders for pathogenic sequences. If someone wants to reconstruct the virus, says Taubenberger, "the technology is available".

Philip Campbell, editor-in-chief of *Nature*, says that although he did not seek advice on whether to publish the work, he has done so for previous flu-virulence and pathogen genome papers. He says that the benefits clearly outweigh the risks. Donald Kennedy, editor-in-chief of *Science*, agrees about the merits of publication. "I think we are going to depend on this kind of knowledge," he says.

The US National Science Advisory Board for Biosecurity (NSABB) reached a similar conclusion about both studies, after calling an emergency meeting last week to consider the risks. But, concerned about public fears, it asked the authors of both papers to add a passage to the manuscripts stating that the work is important for public health and was conducted safely.

Campbell says he is worried that government agencies will start seeking to be involved in the publishing process. "We are happy to cooperate with the NSABB to consider the principles by which dual-use results can be published responsibly," he says. "But government bureaucracies and committees may push to avoid perceived risks, at the potential expense of benefits to public security."

Taubenberger admits that there can be no absolute guarantee of safety. "We are aware that all technological advances could be misused," he says. "But what we are trying to understand is what happened in nature and how to prevent another pandemic. In this case, nature is the bioterrorist."

from CDC head Julie Gerberding and Anthony Fauci, director of the National Institute of Allergy and Infectious Diseases, based in Bethesda, Maryland.

But the studies have sparked fears among other researchers. "There most definitely is reason for concern," says Richard Ebright, a bacteriologist at Rutgers University in Piscataway, New Jersey, who serves on biosecurity panels. "Tumpey *et al.* have constructed, and provided procedures for others to construct, a virus that represents perhaps the most effective bioweapons agent now known."

"This would be extremely dangerous should it escape, and there is a long history of things escaping," says Barbara Hatch Rosenberg, a molecular biologist and member of the Federation of American Scientists' Working Group on Biological Weapons. "What advantage is so much greater than that risk?"

Ebright agrees that there is a significant risk, "verging on inevitability", of accidental release of the virus into the human population, or of theft by a "disgruntled, disturbed or extremist laboratory employee". And there is the danger that a hostile nation might reconstruct its own version of the virus, he says, pointing out that

ON THE RECORD

“Had the decision been mine, we would not have built the space station we’re building in the orbit we’re building it in.”

NASA administrator Michael Griffin attacks the International Space Station.

“Having a moon is just inherently cool — and it is something that most self-respecting planets have.”

Astronomer Michael Brown talks about the discovery of a moon orbiting the Solar System’s ‘tenth planet’.

Source: USA Today, Caltech

SCORECARD**Dogs**

Labradors with

osteoarthritis are the

focus of a study launched at the University of Liverpool, UK.

Researchers will use magnetic resonance imaging to track disease progression in the dogs.

**Nuclear safety**

Ukrainian officials have recovered 14 pieces of nuclear fuel stolen from the now-defunct Chernobyl power plant. The rods, found in a plastic bag near the plant’s perimeter, had been missing for a decade.

**Archaeology**

Embarrassed Czech archaeologists have found that a statuette thought to represent a fifth-century Persian goddess came from a mould made in 1968 by a local pensioner.

NUMBER CRUNCH

60% of PhD-granting physics departments in the United States report visa problems for foreign students returning after trips abroad.

48% of the physics departments had at least one foreign student who was denied entry or considerably delayed by visa problems.

13% is the fall in the number of first-year enrolments by foreign graduate students in the United States between 2000 and 2004.

Source: American Institute of Physics

One to one: tailoring prescriptions to patients’ genes could help to reduce side effects.



AMANA

Japan jumps towards personalized medicine

TOKYO

Japanese companies say they have built a desktop machine that will allow doctors to assess patients’ DNA from a single drop of blood, and so tailor treatment to an individual’s genes. The machine can deliver results within an hour, they say, and will be on sale for 5 million yen (US\$44,000) by autumn 2006.

Safe dosage, effectiveness and side effects for any given drug vary from patient to patient. And determining which drug and dosage is best for any given individual is a critical challenge facing healthcare specialists.

But the announcement about the Japanese machine on 27 September came just a week after scientists in the United Kingdom spoke out against the hype surrounding personalized medicine. A report produced by the Royal Society warned that prescriptions tailored to a patient’s genes are at least 15 to 20 years away.

The Japanese device was developed by the genomics facility of the Institute of Chemical and Physical Research (RIKEN) in Yokohama, the printing company Toppan in Tokyo, and the Kyoto-based maker of scientific equipment Shimadzu. Shimadzu’s Takaaki Sato, who led Shimadzu’s development efforts, says the key advance is a chip that analyses DNA in a blood sample, thereby bypassing the time-consuming DNA purification steps currently needed.

Although Sato will not give further details, he says that any health worker could use the machine to test a drop of blood for a particular genotype, and get a result in an hour. “Patients do not want to wait a week or even a day for

results before being able to take a medicine, especially if they have an infectious disease,” says Sato.

According to RIKEN and Shimadzu, the machine will first be tested on patients being prescribed one of two medicines: an antibody called irinotecan, which can cause hearing loss in people with a certain mutation in their mitochondrial DNA, and the anticoagulant warfarin, which causes excessive bleeding in some patients.

But there is scepticism over how useful the device will be. David Weatherall of the Weatherall Institute of Molecular Medicine at the University of Oxford, UK, who worked on the Royal Society report, says the metabolism of warfarin is related to at least two genes whose interaction is not understood. Other factors, such as the patient’s age or additional drugs being taken, also need to be considered, he says.

“There is no way around these problems except to test each drug independently in large population studies, and to monitor all these issues over several years,” he says. “There is still a huge gap between the scientists who do this kind of work and its application for practical purposes.”

Sato admits that initially his machine will be most useful for research. But judging from the “unbelievable number of responses” received since announcing the test, he says there is no way that 15 years will pass before doctors are using such devices for day-to-day diagnosis and treatment.

David Cyranoski

Q marks the spot as ancient sculptures yield their origins

Archaeologists have pinpointed the location of a long-sought Maya city in Guatemala. The discovery could bolster scientists' battle against local ranchers, who have been moving into the surrounding national park claiming the land as their own.

For several decades, Maya experts have puzzled over some 30 enigmatic sculptures — most of them carved stone blocks — that came to light around the world but seemed to have originally come from the same place. No one knew where the site was. Archaeologists dubbed it 'Site Q', the Q standing for the Spanish '*qué*', meaning 'what?' or 'which?'.

Last week, a team of archaeologists announced that they had found Site Q. It was, they said, the ancient city known as La Corona in the Petén region of northern Guatemala.

La Corona came to the attention of archaeologists in 1997, when Harvard University researchers Ian Graham and David Stuart visited and documented the ruins. They found sculptures that resembled the Site Q monuments. "I've always been convinced that La Corona is the place," says Stuart, now at the University of Texas at Austin.

The latest finding may convince everyone else. In April, a small team briefly visited La Corona under the direction of Marcello Canuto, an archaeologist at Yale University in New Haven, Connecticut. Canuto found the Site Q connection by chance, while trying to escape the jungle's relentless mosquitoes.

He came across a trench dug by looters years before, and there discovered an elaborately carved stone block still in its original setting. The carvings on the stone matched those of the Site Q sculptures and thus prove that La Corona is Site Q, says Canuto.

Hieroglyphs on the panel date it to the year AD 677 and describe key rulers, religious dates and other events. The panel also describes how La Corona was a key ally of the city of Calakmul; the Maya world at the time was ruled by two cities, Calakmul and Tikal, which were fierce enemies.

Unlike the well-known monuments at Tikal, the recordings of Calakmul's history have mostly eroded away, says Canuto. "If La Corona was a faithful ally, this discovery will add a lot of information about Calakmul's history," he adds.

Meanwhile, David Freidel, another team member and an anthropologist at the Southern Methodist University in Dallas, Texas, is working with Guatemalan researchers and conservation agencies on another aspect of the find. They are trying to get the area designated as an archaeological park that should be protected. La Corona lies in a national park that is meant to be kept pristine. But ranchers threaten the sites more each year, says Freidel. During some seasons, the fires they set to clear land can come within a few kilometres of Maya sites. ■

Alexandra Witze

"The carvings on the stone match those of the Site Q sculptures and thus prove that La Corona is Site Q."

R. MCNAB, WILDLIFE CONSERV. SOC.



Hieroglyphs on a stone panel (left) found at La Corona describe key features of Mayan history.



GORILLAS BRANCH OUT INTO TOOL USE

Wild gorillas seen using walking sticks and plank bridges.

www.nature.com/news

T. BREUER ET AL./PLOS BIOL.

Electric current captures top sperm

SYDNEY

Fertility researchers have devised a way to isolate high-quality sperm from a sample of semen using an electric current. The team hopes that the method, which has resulted in the birth of a healthy baby, will lead to more efficient *in vitro* fertilization (IVF) treatments.

"We've been separating sperm in the same way since the first IVF baby Louise Brown was born, more than 25 years ago," says Steven Fleming, scientific director of the Westmead Infertility Centre in Sydney. "No one has significantly moved the technology forward, until now." Fleming hopes to collaborate with the researchers who developed the technique in clinical trials of the new method.

During IVF it is important to separate the sperm from the rest of the semen as quickly as possible because the semen contains potentially destructive oxidative chemicals. This is normally done by centrifugation, but the multiple steps necessary take up to an hour and the forces involved can damage the sperm. Centrifugation normally selects for higher-quality sperm, as sperm with dense intact heads are separated from the rest. Alternatively, sperm can be selected on the basis of their ability to swim.

But neither method works well when sperm samples have to be taken directly from the

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Current methods to select sperm for IVF are slow and difficult.

testicles. In this case, the resulting sperm aren't motile enough for a swim test and cellular debris makes centrifugation difficult.

The new method, developed by researchers at the University of Newcastle in New South Wales, relies on a technique known as electrophoresis, and the fact that negatively charged sperm have the most intact DNA (C. Ainsworth *et al. Hum. Reprod.* 20, 2261–2270; 2005).

Semen is placed in a chamber within a membrane containing pores through which only the sperm can pass. An electric current draws negatively charged sperm across the

membrane, leaving the others, and any contaminants, behind.

It is not known why negatively charged sperm have the most intact DNA. Researchers think that negatively charged sialic acid, added during the final stages of sperm production, is an indicator that the sperm has successfully assembled and matured. "It's like the cherry on the cake," says John Aitken, one of the researchers who developed the technique. "Everything else has to have gone absolutely normally for the sperm to get to that point."

The new separation process takes only five minutes, is gentle, and works on sperm taken directly from the testicles. It also involves fewer steps than centrifugation, reducing the chance that samples could get mixed up.

Aitken and his colleagues collaborated with experts at an IVF clinic in New South Wales to treat a couple who had been unable to conceive using other methods of sperm separation. "Electrophoresis was tried as a last resort," says Aitken. The couple conceived successfully and a healthy baby was delivered earlier this year. The researchers declined to reveal further details, saying they want to publish the results of the case study, but they plan to seek approval from hospital human ethics review committees to start a clinical trial next year.

Carina Dennis

Australia mooted as dump for world's nuclear waste

SYDNEY

Australia should become an international repository for nuclear waste, according to a former prime minister. The idea has outraged environmentalists, but some scientists are giving it cautious consideration.

Robert (Bob) Hawke, prime minister from 1983 to 1991, made his suggestion to a gathering of graduates in Sydney on 26 September. Australia has an obligation, he says, as one of the world's largest uranium suppliers, to be part of the solution for disposing nuclear waste. "We would be doing a good turn, environmentally, for the world," he told *Nature*.

Overseas nuclear-power users would pay to ship their waste to Australian shores, where it would then be transported to sites within the vast, sparsely populated regions of Western Aus-

tralia or the Northern Territory. Hawke claims that the arrangement would be worth billions to Australia's economy. "It would be an enormous source of income that we could use to address our own environmental problems," he says.

"It's not a far-fetched idea," says nuclear physicist Aidan Byrne, who heads the department of physics at the Australian National University in Canberra. Australia's geological and political stability makes it an attractive site for waste disposal, he says.

Environmentalists disagree. "It's fanciful," says Ben Pearson, an energy campaigner for Greenpeace in Australia, who argues that it would be too dangerous to transport large amounts of nuclear waste around the world. "Ships sink; accidents happen," he says.

Hawke plans to rally further discussion on

the topic. "I want to get a sensible debate going," he says. "I would like the Australian scientific community to put resources into confirming the safest sites."

The proposal will struggle to get political support; the federal and state governments cannot even agree on where to store the nation's own small amount of low-level nuclear waste. And several states have legislation that bans the import of nuclear waste.

But Byrne thinks attitudes may be shifting. "If Hawke had suggested this a year or so ago, it would have been seen as ridiculous. But the nuclear debate has come a long way," he adds, "if Australia is to be part of the nuclear cycle as a supplier, then we need to be thinking about waste disposal as well."

Carina Dennis

Physics prize puts spotlight on optics

M. OLDFIELD, SCUBAZOO/SPL

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Work by this year's physics laureates led to a better understanding of the nature of light.

Three researchers who applied quantum theory to light, and built devices that are now providing the best-ever measurement of fundamental constants, have been awarded this year's Nobel Prize in Physics.

Half the award goes to the theoretician who laid the groundwork for the advances: Roy Glauber. A professor of physics at Harvard University since 1976, Glauber is a former member of the Manhattan Project — the effort that led to the development of atomic weapons during the Second World War.

John Hall of the University of Colorado in Boulder and Theodor Hänsch of Ludwig Maximilians University in Munich, Germany, share the other half for developing techniques to measure the frequency of light emitted by atoms and molecules.

Hänsch was busy at his university, packing for a flight to the United States, when he heard the news. Half an hour later, he was trying to deal with the attentions of the hundred or so reporters who had arrived at his office, eager to know everything about his life. He says he never expected to win the award. "I'm overwhelmed," he says. "I haven't absorbed it yet."



A. ABBOTT

Theodor Hänsch is besieged by reporters after winning a share in this year's physics Nobel.

Hänsch and Hall's work is rooted in two papers published by Glauber in 1963, which built on the excitement in the physics community generated by the development of lasers in the 1950s. The papers focused on the working of the devices that are used to measure photons of light from lasers and other sources. Glauber showed that normal statistics failed to describe the interaction between photon and detector; only an understanding of the

California prepares to roll out stem-cell funding

SAN FRANCISCO

The funds for California's stem-cell research initiative may finally start flowing. The California Institute for Regenerative Medicine (CIRM) has been unable to use the US\$3-billion that it was granted by California voters nearly a year ago because of ongoing legal wrangles. But on 3 October, state officials began the process of selling up to \$55 million in bond anticipation notes for the first grants.

Sixteen research institutions are anxiously awaiting the first \$12.5 million, awarded last month as training grants for 170 graduate students, fellows and postdocs over the next year. And last weekend, leading stem-cell researchers met in San Francisco to discuss what to do with the money.

The CIRM was formed to conduct studies that the US government will not fund because of restrictions laid out by President George W. Bush. The grants are part of the institute's unprecedented effort to create a

huge research programme from scratch, while fighting off legal and political challenges from those who oppose studies involving human embryos (see *Nature* 434, 694–696; 2005).

At the San Francisco meeting, prominent stem-cell researchers, including those from Canada and Sweden, joined US colleagues to help the CIRM chart research paths. The nearly 150 scientists in attendance called for basic research with an eye to advancing to clinical trials as quickly as possible. Suggested topics include how stem cells differentiate, the immunological barriers to stem-cell transplantation, and creating new imaging techniques to track stem cells transplanted into humans.

Among those at the meeting, stem-cell researcher Andras Nagy of Mount Sinai Hospital in Toronto disclosed that he and his colleagues had developed an embryonic stem-cell line from dogs, which they hope may shed light on species differences and so

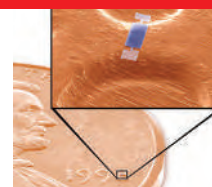
help researchers apply the results of studies in mice to humans.

Legal challenges have slowed the CIRM's plan to issue about \$250 million annually in grants over the next decade. Three Sacramento-area groups — the People's Advocate, the National Tax Limitation Foundation and the California Family Bioethics Council — filed lawsuits earlier this year in the state court. They claimed that Proposition 71, which California voters overwhelmingly approved last November to create the CIRM, was technically flawed.

Two of these suits were combined into a single action at a superior court in August; a judge is to issue a crucial ruling on 17 November, when state attorneys will request that the lawsuits be dismissed.

If they are dismissed, the state will still have to wait about a year for the appeals process to finish before it can start selling bonds to fund the full annual research amounts. Meanwhile, on 3 November the state announced plans to sell bond anticipation notes to philanthropic

"Despite the ongoing legal and political battles, grant awardees and universities are charging ahead."



METAL DETECTORS GET SHRINKING FEELING
Tiny prototype promises better military reconnaissance.
www.nature.com/nature

ILLUSTRATION: LUCENT TECHNOLOGIES BELL LABS

quantum nature of the device could explain it.

Such breakthroughs turned Glauber into a major player in the emerging field of quantum optics, which applies established quantum theory to light. The fruits of the discipline were a better understanding of lasers and the process by which excited atoms and molecules emit photons of light. Hall and Hänsch independently used such results to build optical combs, which are laser devices that can measure the frequency of light sources with great precision.

Fields as diverse as navigation and cosmology are benefiting from optical combs, which are being used to develop a new generation of optical clocks — potentially capable of a precision of 1 part in 10^{18} . Researchers studying the fine-structure constant, which determines the strength of the interaction between light and matter, are using the clocks to study whether the constant changes slightly with time. The devices could also lead to a redefinition of the second, and help to improve the precision of the navigation signals emitted by global positioning systems. ■

Jim Giles

Additional reporting by Alison Abbott

The Nobel Prize in Chemistry was announced after Nature went to press. For coverage see

► www.nature.com/news

organizations that could serve as a financial bridge for the first \$55 million.

Despite the ongoing legal and political battles, CIRM officials, grant awardees and universities are charging ahead. Even though no money is yet available, the University of Southern California in Los Angeles will soon announce plans for a large new building with a wing devoted to stem-cell research, says Francis Markland, associate dean of scientific affairs, the construction of which is to begin early next year. The university is working to recruit six to eight professors to positions there.

Elsewhere, young researchers and senior faculty members have put career decisions on hold as they consider moving to California institutes in search of CIRM research funds. "This does cause some anxiety," says physician Robert Mahley, president of the J. David Gladstone Institutes in San Francisco, which is to receive \$2.4 million in training grants over three years. "But we are prepared to wait it out." ■

Rex Dalton

Gut feeling secures medical Nobel for Australian doctors

Barry Marshall and Robin Warren have won this year's Nobel Prize in Medicine or Physiology for discovering that most stomach ulcers are caused by the bacterium *Helicobacter pylori*. Despite original resistance to the findings, their work at the Royal Perth Hospital has revolutionized the treatment of gastric disease.

Plain-speaking Barry Marshall has long been a folk hero in his native Australia. But in the years after his 1982 discovery, he was dismissed as an upstart who was pushing a hypothesis that had no credibility. That pushiness, combined with dogged determination and sharp insight, kept alive the heretical idea that gastric and duodenal ulcers could be caused by a bacterial infection.

At the time, ulcers were treated with drugs that reduced acid secretion in the stomach. The drugs worked, so acidity was assumed to cause ulcers. But pathologist Warren had noticed spiral-shaped bacteria in biopsies from ulcerous stomachs, and that these were always associated with inflammation. He was convinced that the bacteria were linked to the ulcers.

He recruited a young medical intern — Marshall — to isolate and grow the bacteria in culture. The bacteria looked like *Campylobacter*, a newly discovered family known to cause gut infection in poultry. But Marshall's initial attempts in 1982 failed — until Easter, when culture plates were accidentally left over the four-day break. It turned out that the bacteria grew extremely slowly, and earlier attempts had simply been abandoned too soon. The bacteria were then shown not to be *Campylobacter*, but an entirely new genus, named *Helicobacter*.

Marshall and Warren went on to show that patients with ulcers can be treated with antibiotics. Unlike patients given acid-suppressing drugs, their ulcers do not return.

But gastroenterologists resisted the idea. Francis Mégraud, a bacteriologist at the Victor Segalen University in Bordeaux, France, remembers attending the 1988 meeting of the American Gastroenterological Association in New Orleans and hearing outraged physicians. "They seemed insulted, saying, 'we are being asked to treat stomach ulcers with antibiotics, as if it were gonorrhoea!'" he says. "It was hard for them to accept that the disease could be a simple infection."

Drug companies that profited from the anti-ulcer drug market were also actively resistant, says Mégraud, who is secretary of the Euro-

pean *Helicobacter Pylori* Study Group. Even some bacteriologists were suspicious — the stomach had long been assumed too acidic to host bacteria.

In frustration, Marshall did the ultimate cause-and-effect experiment. He swallowed a solution containing the bacteria, and promptly came down with an aggressive attack of the sort of gastritis that leads to ulcers. "My colleagues were alarmed, and so was my wife," he recalls.

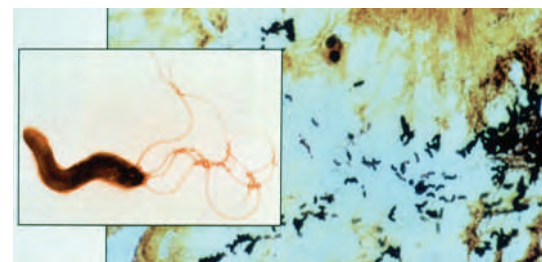
Marshall's forthright attacks on doubters did little to soften critics. Their prejudices were deepened by his youth, and the fact that Perth had no strong academic reputation. "Fortunately, I'm very thick-skinned," he says. "There was also an advantage to being isolated in Perth. I don't think I realized just how heavy the opposition was."

His untiring advocacy, and further research with Warren, subsequently repeated and extended around the world, eventually won the day. In 1991, a meeting of the Centers for Disease Control and Prevention in Atlanta, Georgia, formally declared the link between *H. pylori* and gastric disease.

It is now accepted that most gastric ulcers are caused by *H. pylori*. The bacterium is usually acquired in childhood, being transferred through faeces or vomit between family members. It then lies dormant until adulthood. Untreated cases can lead to gastric cancer. ■

Alison Abbott

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS



Robin Warren (left) and Barry Marshall found that *Helicobacter pylori* (bottom) causes stomach ulcers.

W. EDWARDS/REUTERS; UNIV. VIRGINIA/AP

Conservationists condemn changes to protection act

The US House of Representatives has approved a major overhaul of the Endangered Species Act, the 1973 environmental law that protects plants and animals threatened with extinction.

The bill, which passed on 29 September, includes such sweeping changes as paying property owners who cannot develop their land because it is home to an endangered species. It would end the designation of specially protected 'critical habitat' for endangered species, and would considerably toughen the requirements for data taken into account when making decisions about a species' fate.

Under the proposal, all data must be approved by the Secretary of the Interior, a move that critics say will ill-advisedly put decisions of scientific value into the hands of a political appointee. Environmental groups are bemoaning the bill's passage. Representative Richard Pombo (Republican, California), who introduced it, says the changes will improve the scientific basis for decision-making and refocus the act on recovering endangered species.

The bill will not become law unless the

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Proposed changes to the Endangered Species Act may affect animals such as the black-footed ferret.

Senate passes a similar bill, which is not expected in the near future. Senators led by Lincoln Chafee (Republican, Rhode Island) are working on less radical reforms to the act.

Head of drugs watchdog drops some cancer duties

Just days after being appointed interim chief of the US Food and Drug Administration, Andrew von Eschenbach has temporarily given up his day-to-day duties as director of the National Cancer Institute. He has also excused himself from many matters

involving cancer drugs that might come before the agency (see page 789).

The dual appointment of von Eschenbach had come under fire from ethicists. They said that one person could not effectively run two multibillion-dollar federal agencies, and that he faced conflict-of-interest problems (see *Nature* 437, 606; 2005).

Germany's prodigal sons tell it to buck up funding

Young German scientists working in the United States have sent an open letter to their research ministry asking it to improve conditions at home. Among other things, they demand more money for research, more transparency in recruitment processes, and a tenure-track system for young scientists, as is common in the United States.

Drafted by the German Scholars Organization, the letter argues that young scientists do not want to return to Germany because of the obstacles there to conducting research. "We need reforms," says Michael Koeris, one of the letter's authors and a biologist at the Massachusetts Institute of Technology in Cambridge.

The 11 researchers who drafted the letter are supported by more than 350 colleagues

— almost all German scientists working in the United States, including Nobel laureates Wolfgang Ketterle and Herbert Kroemer.

Japan's fast-breeder reactor gets green light

Japan should aim to commercialize its prototype fast-breeder nuclear reactor by 2050, says a panel convened by the

country's Atomic Energy Commission.

The Monju reactor in Tsuruga, central Japan, has been shut down since 1995, when it began leaking liquid sodium just a few months after it started to operate. Japan plans to restart the fast-breeder reactor, which produces more fissile material than it consumes, in 2008.

Officials in Japan's cabinet office argue that the reactor is now safe, and that the target date will motivate researchers to

accelerate research and development. Many of the country's nuclear power plants will need to be replaced by 2050, and the Monju reactor, which it is hoped will be more cost-effective, could help to make replacements more affordable, the atomic-energy panel reported on 29 September.

Meanwhile, France's Atomic Energy Commission has reaffirmed that it is interested in jointly using the Monju reactor, according to a Japanese government official.

Cassini gets close to Saturn's battered satellite

It looks like the strangest moon in the Solar System. When the Cassini probe passed just 500 kilometres above Saturn's satellite Hyperion on 26 September, it revealed a landscape covered in deep craters with sharp rims (see picture).

"The surface looks weird," says Tilmann Denk, a planetary scientist at the Free University in Berlin. "It looks so different from anything we've seen before, like a sponge you would use in your bathtub."

Hyperion has an unusual elongated shape, being 360 kilometres long but just 250 kilometres wide, and it tumbles around Saturn chaotically. Because other objects of a similar size are much more spherical, astronomers think that Hyperion may be a fragment of a larger object that was smashed apart by a collision.



Correction

Magnets and equipment supplied by Oxford Instruments, as described in "Fatal attraction" (*Nature* **436**, 624–625; 2005), are used for nuclear magnetic resonance spectroscopy not magnetic resonance imaging, as stated in the article. The world's first superconducting magnet was also made three years after the company's foundation, not beforehand.

Clarification

In our News story on using fMRI lie detection to uncover criminals (*Nature* **437**, 457; 2005), we stated that Daniel Langleben and his colleagues could detect lies from truth with 99% accuracy. We wish to clarify that this was the accuracy for individual tests — the more likely average success rate for practical purposes is about 88%.



Flight of the navigators

The Arctic is a unique testing ground for studying how birds navigate long distances.

Jane Qiu catches up with an expedition to unravel the signals that help birds on their migrations.

The captain of an ice-breaker has few problems navigating during the polar summer. The 24-hour sunlight provides constant illumination of the surroundings. When storms whip up, the captain can turn to a suite of global-positioning devices, which pinpoint the ship's location to a matter of metres. Yet all this electronic sophistication is put to shame by the migratory birds wheeling overhead, which navigate thousands of kilometres using nothing more than what is in their heads.

This summer, the ice-breaker *Oden* tracked migrating birds as they left the Arctic through the Bering Strait, the narrow waterway between Alaska and Siberia. More than 50 polar scientists had gathered as part of a wide-ranging project called Beringia 2005. Among them were Thomas Alerstam of Lund University in Sweden and his research team. They were there to shed light on one of ornithology's greatest mysteries: how do birds navigate during their annual migration?

Thousands of birds gather from all over the world to breed and rear their young in the summer Arctic, where the landscape is temporarily rich with thousands of plants and insect species. Birds such as the Arctic tern (*Sterna paradisaea*) fly 18,000 kilometres across featureless oceans to get there, and navigational skills are crucial. Drifting off course could cause the birds to miss the land and succumb to exhaustion. When summer wanes, they must once again find their way back,

heading as far as South America, New Zealand or even the Antarctic.

The precise nature of their navigational gift has fascinated scientists for hundreds of years¹. In the early nineteenth century, studies in basic magnetism, along with expeditions to the North and South Poles, inspired scientists to propose that birds are guided by an inner magnetic sense, like a compass needle. Nearly two centuries later, this seemingly bizarre idea was supported by the discovery of magnetite, or iron oxide, in the brains of some bird species. The magnetite is thought to act as a tiny compass².

Clued up

But birds use a number of different navigational cues. As well as Earth's magnetic field, these include the landscape, and the positions of the Sun and the stars. In controlled laboratory experiments, scientists have demonstrated that a number of bird species can use these cues and others, separately or in combination. However, no one knows how the birds put them together to find their way. "It is a different matter in the wild," says Alerstam. "Which cues do they use then? Are some cues more important than others? We have no answer to these questions yet."

To navigate, birds need to know in which direction they are heading. A magnetic compass sense can tell them; so can the position of the Sun, as long as they know the time of day. But direction alone cannot keep a bird on

course, says Sönke Johnsen, a biologist at Duke University in Durham, North Carolina. "A compass sense is often not enough to guide an animal to a specific destination or to steer reliably along a long and complex migratory route," he says. Birds must also be able to determine their position relative to a destination and continuously recalculate their heading.

Just three parameters of Earth's magnetic field may be enough to guide the birds on their journeys, researchers think. The first, the strength of the magnetic field, varies with latitude. The second, dip angle, is the angle between the magnetic field line at a given location and Earth's surface; this angle is 0° at the equator and 90° at the north and south magnetic poles. The third parameter, magnetic variation, marks the difference between directions towards the geographic and magnetic poles, and varies depending on location on Earth's surface.

Exactly how birds derive their direction and position from this information isn't entirely clear. The problem is further complicated at high latitudes, where Earth's magnetic field lines converge. "As the dip angle approaches 90°, the magnetic field lines go almost straight into the Earth, leaving very little of the horizontal information that is necessary for orientation," says Alerstam. As birds get closer to the north magnetic pole, which lies 1,400 km from its geographic counterpart, magnetic variation increases. And magnetic minerals in the Arctic Ocean or the tundra can give rise to



T. ALERSTAM

Ocean cruisers: flocks of birds are tracked by the ice-breaker *Oden* (left) and its radar (right) as they migrate to the richness of the Arctic tundra (above).

anomalies in the magnetic field in the region. Daily variations in the magnetic field are also more erratic, because of the way that solar radiation affects the magnetic poles.

The 24-hour daylight of the polar summer means that the birds cannot even use the stars or daily rhythms of the Sun as a pointer. But the very complexity of the Arctic makes it an attractive location for testing birds' navigational strategies. "The only way to find out how birds cope in such complicated geomagnetic regions is to go out there and take a look," says Alerstam.

Bearing up

Two months into this summer's expedition, Alerstam wasn't certain that his team would find any answers. At 4:00 a.m. on 14 August, *Oden* was steering slowly through thick ice towards Wrangel Island off Siberia. The ship swayed heavily from side to side and the ice was so dense that it had to reverse slightly to gain enough momentum to advance another 50 metres. "We can only hope that the equipment will survive such harsh treatment," Alerstam wrote in his notes.

After a few hours, the ice thinned and *Oden* started to sail smoothly. In the ship's operation room, Alerstam was absorbed in watching hundreds of echoes appearing on the radar screens. Conditions were perfect after several days of rain, and *Oden's* radar equipment had picked up echoes of large flocks of migratory birds heading out of the Arctic. Most of them were travelling east and south at high altitudes, 2,000 to 3,000 metres up. One echo even came from an altitude of 4,800 metres — a record height for the trip.

Radar tracking allows scientists to follow individual birds or flocks for up to 15 kilome-

tres. By the end of the trip, Alerstam and his team had recorded nearly 600 such tracks. Based on these echo recordings, they calculated flock densities as well as the speed, direction and altitude of the migration. The motion of helium-filled balloons released from *Oden* was also tracked, and this information on wind patterns allowed the team to calculate the headings and trajectories of the birds themselves.

The different cues that birds might use for navigation generate very different trajectories, particularly at high latitudes. So researchers

"Expeditions such as Alerstam's are pushing bird-navigation studies from a steady state to a new level."

— Martin Wikelski

can calculate specific paths for each cue and compare them with where the flocks actually go, revealing which cues are most important to the birds. Alerstam calculated several possible routes for the species they wanted to track.

He predicted that, if the birds navigate mainly using a magnetic compass, they would fly northeast from the expedition site towards high Arctic Canada.

Predicting paths for birds that are using a Sun compass is more complicated because the directional information depends on the bird's sense of time. At high latitudes, the distances between longitudes are so small that migratory birds may fly across three or four lines of longitude in a single day. Either the birds' internal clocks adjust to the local time as they go, or they remain constantly jet-lagged. Alerstam calculated that birds following the Sun and not

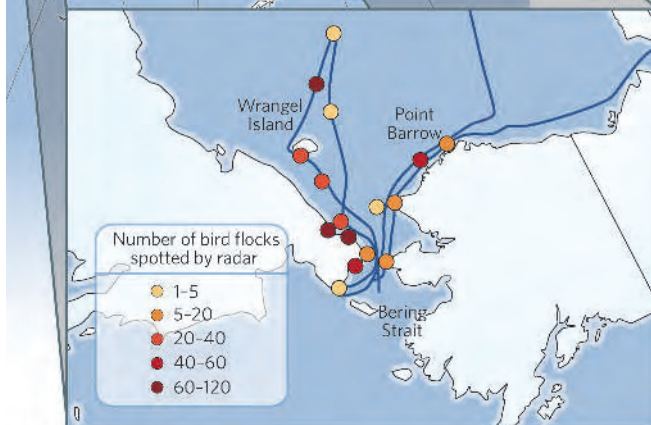
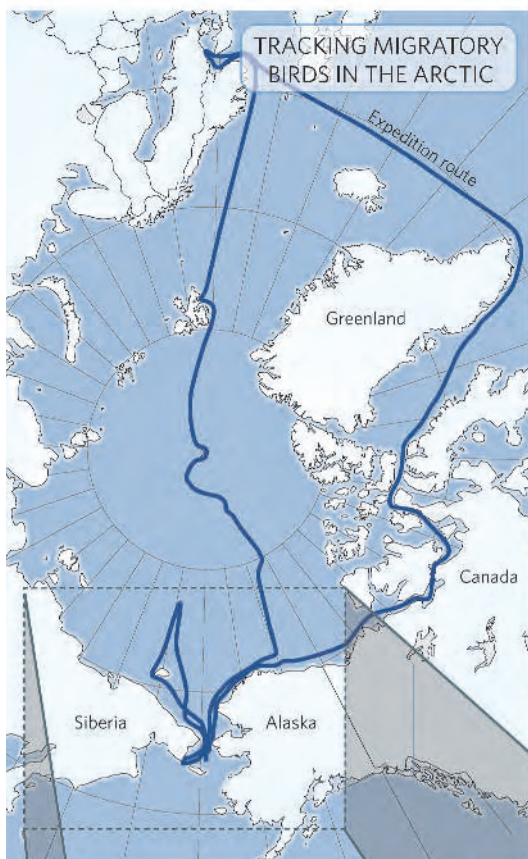
adjusting to local time would curve to the southeast, passing through the Bering Strait and western and northern Alaska. Birds that fly using a Sun compass and correct it for the shifting time zones might be expected to fly a route somewhere between the Bering Strait and the magnetic route.

In 1994, Alerstam went on an expedition to the Russian Arctic and discovered that most birds in western Siberia flew towards the east at the end of the breeding season, a puzzling observation as the birds' destination was south. It turned out that their migratory paths were most consistent with a Sun-compass route without time correction, and Alerstam inferred that the birds were using the Sun to navigate while experiencing constant jet lag³. So on the *Oden* trip, says Alerstam, "we expected to see massive migrant flow — mostly on southeast courses — over the Bering Strait, western and northern Alaska, and northwesternmost Canada." That is exactly what his team found (see map, overleaf).

Coarse corrections

One curious feature of the strategy that uses the Sun without time correction is that it gives trajectories that approximate a 'great circle' route. A great circle is the shortest arc connecting two points on the globe. Following a great circle route is quicker, but requires the traveller to change compass course continuously; navigating with a constant magnetic compass course is easier, but results in much longer routes. Pilots and sailors regularly follow great circle routes with the help of complex geometric computing. And it seems that birds have found a way to do this too.

"It makes sense as this saves energy, which is important given that they have to fly



Spotting birds through telescopes helps to identify flocks picked up by radar.

thousands of miles," says Alerstam.

How have birds learned this navigational trick? For most animals that travel in east-west directions, changes in the time of sunset or sunrise reset their internal clock — but this process takes a few days. If it also takes days in Arctic birds, then as they migrate long distances towards the east or west they will be constantly out of phase with the local time, and misread the Sun as a result. Not knowing that they are out of sync, they will end up flying along the energy-saving great circle routes. "Maybe it is a lucky coincidence," says Alerstam.

Another possibility is that the birds can correct for the changing time zones, but actively suppress this adjustment. According to the late Eberhard Gwinner of the Max Planck Institute for Ornithology in Andechs, Germany, migratory birds in the Arctic should be able to adjust their internal clock very swiftly⁴. In the Arctic

Of course birds must have some sense of time to find their direction, even if it is then slightly offset by jet lag. "There are other natural cues the birds rely on, such as the colour and polarization patterns of light," says Michaela Hau of Princeton University in New Jersey.

Once the birds find their way to lower latitudes, points out Alerstam, the conditions for navigation become much less extreme. He suggests that they may then use different strategies to find their way.

Are we nearly there yet?

One thing is clear from the past 30 years of research: birds use a delicate combination of cues to create a sophisticated backup system. If one cue is taken away from them, they will use the next down the line. And the cues that birds use and the way they navigate depends crucially on geographic location and weather conditions.

"At the end of the day, both fieldwork and laboratory experiments are necessary for understanding bird orientation. And they should always go hand in hand," says Martin Wikelski of Princeton University. "Expeditions such as Alerstam's represent one of those points where we are pushing things from a steady state to the next level," he says.

In another such push, Wikelski is working on an international initiative to study small-animal migration around the world. This project will use unmanned aerial vehicles or even a specialized radio receiver in low Earth orbit. Either approach would allow researchers to map the global migratory patterns of birds and insects carrying embedded miniature radio transmitters.

"To track individual birds around the globe will completely revolutionize avian research," says Alerstam. Such studies will not only shed light on bird navigation, but also have implications for preventing or containing animal-based epidemics such as avian influenza. The study of migrating birds may thus have a bearing on much wider questions than just how and why they fly. ■

Jane Qiu is an editor for *Nature Reviews Neuroscience*.

1. Johnsen, S. & Lohmann, K. J. *Nature Rev. Neurosci.* **6**, 703-712 (2005).
2. Walcott, C., Gould, J. L. & Kirschvink, J. L. *Science* **205**, 1027-1029 (1979).
3. Alerstam, T. & Gudmundsson, G. A. *Proc. R. Soc. Lond. B* **266**, 2499-2505 (1999).
4. Gwinner, E. & Brandstätter R. *Phil. Trans. R. Soc. Lond. B* **356**, 1801-1810 (2001).

The friction that arises when a scientific society aims both to serve its members and stay commercially competitive is generating heat within the American Chemical Society.

Emma Marris takes the society's temperature.



Chemical reaction

The American Chemical Society (ACS) is the world's largest scientific society. Composed of more than 158,000 chemists from industry and academia, and employing some 2,000 members of staff, the society's assets total \$1 billion. It publishes journals, holds meetings, provides career services, educates the public about chemistry, rubs shoulders with lawmakers, trains teachers, gives out grants and scholarships, and even sells insurance.

The society owes most of its wealth to its two 'information services' divisions — the publications arm and the Chemical Abstracts Service (CAS), a rich database of chemical information and literature. Together, in 2004, these divisions made about \$340 million — 82% of the

society's revenue — and accounted for \$300 million (74%) of its expenditure. Over the past five years, the society has seen its revenue and expenditure grow steadily (see chart, overleaf).

Although the ACS is a non-profit organization, the information-services divisions are increasingly being run like businesses. Any net revenue is naturally fed back into the society's other activities, but the business-like attitude is making some ACS members uneasy. A small but vocal group of critics fears that business priorities are supplanting the goal laid out in the society's charter: "to encourage in the broadest and most liberal manner the advancement of chemistry and all its branches". This is the heart of the matter, for the ACS, or for any society with substantial

revenues: can a non-profit organization always afford to put its members' interests first?

An ongoing dispute between the ACS and the US National Institutes of Health (NIH) reflects some of the problems. The NIH has recently unveiled a freely accessible database called PubChem, which provides information on the biological activity of small molecules. The ACS sees this as unfair competition to the fee-based CAS because it is taxpayer-funded, and the society wants the database restricted to molecules that have been screened by NIH centres. A few ACS members argue that the society is being unduly aggressive in protecting CAS and ought not to be challenging the scope of a database that could be a useful and free resource for chemists. For the record, *Nature's* sister journal *Nature Chemical Biology* links all of its articles to PubChem.

"I am growing increasingly upset with their direction," says Chris Reed, an inorganic chemist at the University of California, Riverside, and one of the more outspoken critics of the ACS. "They have a culture of a for-profit corporation."

The ACS says that it is run by its members, and that the 2,000 staff ultimately follow the will of elected officers, including the democratically elected 16-member board of directors, who sit at the top of a complex system of 'governance'. Between 13,000 and 20,000 members serve on the society's innumerable boards and committees, although many boards meet only

K. SIMPSON



Members' club: President William Carroll (right) and executive director Madeleine Jacobs (centre) describe the American Chemical Society (left) as a largely democratic organization.



REPRODUCED WITH PERMISSION FROM CHEM. ENG. NEWS 83(38), 44 (2005) © ACS

ACS

a few times a year and some have limited advisory power. Consequently, the hired staff make most of the operational decisions.

"We have a very democratic organization," says William Carroll of Occidental Chemical and president of the ACS. "I think the organization, for one as big as it is, is pretty responsive." Staff and volunteers in governance believe that the ACS can run a competitive business while putting its members' interests first. "The staff here serves governance," says Madeleine Jacobs, executive director of the society and former editor-in-chief of its weekly magazine, *Chemical and Engineering News*.

But membership dissatisfaction bubbled to the surface on 29 August at the society's national meeting in Washington DC during an open session for the CAS and publications divisions. ACS member David Spellmeyer of the IBM Almaden Research Center in San Jose, California, had distributed yellow flyers encouraging people to attend the meeting if they wanted to know more about the PubChem debate. But at the meeting, the director of CAS, Robert Massie, twice declined to discuss PubChem, each time suggesting that interested members talk privately with Brian Dougherty, senior ACS adviser on PubChem who sat at the very back of the room. When he passed over the topic for a second time, saying, "As I said, Brian is here. I suggest you talk to him directly," a substantial chunk of the meeting got up and left the room.

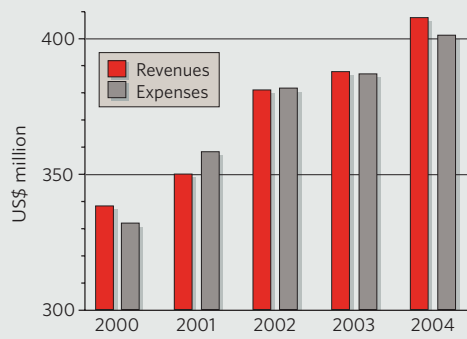
Heated argument

Outside in the corridor, Dougherty found himself surrounded by a knot of angry chemists clutching yellow flyers. Dougherty was conciliatory, but the members were obviously frustrated. "This is offensive," said one.

Dougherty and other staff had first seen the flyers that morning. They had not planned to mention PubChem at the meeting and they decided to stick to the programme, which featured updates such as the introduction of CAS for BlackBerry wireless Internet devices. "They are normally one-hour meetings," says Dougherty. "Was it appropriate for us to turn that into an open forum on PubChem?" He says he suspects that both Jacobs and Carroll would have liked to be present at any such debate. "That was the best decision at the time, and in hindsight, I think we did the right thing," he adds.

Not all ACS members are happy with that decision. Two, including Spellmeyer, resigned in protest from an advisory committee on CAS following the incident, which came just a day after the committee passed a resolution calling for more open dialogue on PubChem between the ACS and its members. One CAS committee member, who preferred not to be named, says: "The ACS had made some choices about what, where and how to communicate. We would like them to rethink those choices. We felt the membership was not well informed."

FINANCES AT THE AMERICAN CHEMICAL SOCIETY



Last year's ACS president, Charles Casey of the University of Wisconsin, Madison, who serves on the board of directors, admits that the society has not communicated well on the PubChem issue. "I think there is an effort now to do better," he says. "I think this is an isolated problem, and one that will get corrected."

Dougherty says the staff has taken the CAS committee's resolution to heart. "We have our orders to improve communication. There is not an action plan, with dates attached, but it will happen," he says.

As dramatic as the scene in Washington was, the crowd in the hallway numbered fewer than a dozen chemists. An ACS survey that randomly polled some 3,500 members by e-mail in 2004 revealed only 5% who said they were not satisfied with the organization. Most approve of the ACS.

Still, as Jacobs says: "There is a small group of people who feel passionately about their issues. Those individuals have real concerns and we have been trying to address them." But policy decisions on questions such as PubChem have to be made expeditiously by the staff, she says. "You can't, for every single issue, poll 158,000 members."

Abstract art

According to Jacobs, CAS is the one thing, besides the society's sheer size, that makes the ACS unique. This operation sells various tools that can bring up a body of research on a chemical with a few keystrokes. "That's an anomaly of the ACS, but also our strength. No other professional society in the world has a publications arm like CAS," says Jacobs.

Based in Columbus, Ohio, CAS is 100 years old, and has become essential for chemists: search tools that mine the core database are purchased by virtually every chemistry com-

pany or research university. Massie, who has a background in management consultancy, is the first non-chemist to head the division. "Bob Massie was definitely hired to move CAS into the modern age, and he has succeeded," says Jacobs. "Its current success is most assuredly due to his leadership."

Under his management, CAS has not been shy of protecting its business interests. In 2002, it sued a company called Leadscope in Columbus, Ohio, founded by three former employees, two of whom are chemists and ACS members. The ACS said that the three were selling a software program too similar to one they had worked on for the society, where they had signed agreements giving ACS the rights to their creations. That suit is on hold, and Jacobs says that she expects it will be settled out of court.

Leadscope chief executive Loftus Lucas said in documents submitted as part of the lawsuit that the ACS timed its suit to scare off a potential financier. "They sued us in a rushed submission, one week before Leadscope was doing a last round of financing," he says. "They elected, apparently, to try to put us out of business."

The society denies this. In a statement it said: "ACS brought suit against the defendants for the reasons stated in the complaint, including breach of employment agreement and misappropriation of intellectual property belonging to CAS. These are the only reasons for the lawsuit, and any other allegations are incorrect."

Scholarly debate

In December 2004, the ACS also launched a suit against Internet search engine Google over trademark infringement for its use of the word 'scholar'. The ACS alleges that the new academic search engine Google Scholar is trading off the reputation built up by the society's popular database search tool, SciFinder Scholar.

And, most controversially, CAS has disputed the scope of the NIH's free database, PubChem. But the NIH sees the two databases as complementary tools and doesn't want to restrict its operations. Negotiations continue between the parties, and visits to lawmakers have been made by both sides. "It takes something pretty dramatic for us to go out to Capitol Hill and not talk about science but about something that would affect our operation," says Dougherty. He predicts that the disagreement will be resolved within a year.

Some ACS members who are also fans of PubChem feel that their voices have been lost in this debate. Steve Heller, who lives in Silver Spring, Maryland, is part of an e-mail list-server community that is a source of lively discussion on this issue. Heller is a retired chemist and ACS member who also serves on an NIH advisory board on PubChem. "It seems as if those members of the ACS who see and know what is going on — and it is not a very large number — are very upset that the management and staff are taking a position without any consultation with the member-

"There is a small group of people who feel passionately about their issues. Those individuals have real concerns and we have been trying to address them."

— Madeleine Jacobs



Some members of the American Chemical Society see the free-to-use PubChem database as a complementary tool to the society's products.

ship or discussion with experts in the field, and doing things that are not in the interest of their members, who want this for free," he says.

"It just seems to me like such a conflict of interest with their own mission," agrees Heather Joseph, executive director of the Washington-based Scholarly Publishing and Academic Resources Coalition, an alliance of libraries that supports open access to the scientific literature. "For science to be conducted effectively, open access to data has to be part of the landscape."

Steady state

Massie told *Nature* he did not have time to be interviewed for this article, but he and other ACS staff and officers are on record as saying that keeping CAS thriving and in the hands of the ACS is the best thing for science. Jacobs argues that for the NIH "to duplicate what the private sector does well, does not seem to me to be a good use of taxpayer money".

Things seem calmer in the publications division, which was reorganized along more commercial lines in the 1990s. "We weren't competing," recalls Ned Heindel, a past ACS president at Lehigh University in Bethlehem, Pennsylvania. "Eventually it was decided that management of the publications should be done by true professionals."

In 1996, oversight moved from a committee of volunteers that met biannually to a Governing Board for Publishing, including four compensated publishing experts. Robert Bovenschulte, a veteran commercial publisher, was hired to run the division. "We are now positioned to respond to the market and move page allocations around," says Heindel.

The ACS has since expanded the number

of titles and done quite well in a field where everyone is feeling the pinch of shrinking library budgets, increasing costs, ballooning submissions, and the parallel rises of the Internet and open-access publications.

In response to the proponents of open access, some of whom argue that if the society's objective is to promote the field of chemistry it ought to give all its journal articles away for free, Bovenschulte replies that the ACS has to charge to keep the publications division strong and dynamic. "In addition, the society does require that both CAS and publications generate a small net contribution or surplus, and those monies support the good works that the society does," he says. "They give me targets and I have to make them." Bovenschulte adds that there have been occasions when business concerns have been trumped by members' interests, but he said he preferred not to name them for fear of reviving old squabbles.

Most critics of the ACS believe that their best shot at changing the society is to find a way to

penetrate the committees and boards that govern it — for example, by electing like-minded individuals to the board of directors, which hires executive staff. Or they could aim for the Governing Board for Publishing, which covers CAS as well as publications. In addition to its four publishing experts, the chair of the board of directors, Jacobs, Massie and Bovenschulte, the board has slots for two ACS members, as long as they have executive-level experience.

Reed hopes to gather together the disgruntled into an e-mail list, and eventually into a slate of candidates for various ACS offices to run on a reform platform. "I think it is going to require a grassroots movement," he says. "I don't see how members' complaints, letters to the editor, or e-mails to the governing board are changing anything."

Many of the critics remain deeply engaged with the society. They attend national meetings and vote in elections. And they generally like the ACS. "On the whole, the society brings a great deal of value to its members," says Spellmeyer. Heller feels the same way. "There are benefits to being a member and being part of a community," he says.

Heller and others hope that they can change the society from the inside. "If you decide you aren't happy with the current governor or mayor, you don't move out of the city or state, you wait for the next election," he says. This sentiment was echoed by Jacobs in a statement released on 26 September: "I firmly believe on PubChem and other issues, it's much more productive to work through our Society, than outside it." On that, at least, there is some agreement.

Emma Marris is a correspondent for *Nature* based in Washington DC.



"I am growing increasingly upset with the ACS's direction. It has a culture of a for-profit corporation."
— Chris Reed

Paper chase

Thousands of patients are queueing to be treated by Hongyun Huang at his Beijing clinic. But no Western journal editor seems willing to publish his research. **David Cyranoski** talks to the neurosurgeon whose global reputation among the ailing hasn't swayed his peers.

Renowned German artist Jörg Immendorff is used to causing a stir. But when he flew to China this March to receive an unconventional treatment for his neurological disease, he provoked uproar at home. His doctor scolded him for trying a therapy that was unproven. And newspapers warned of the possible rise of medical tourism to China based on false hopes.

Immendorff was undeterred. He suffers from ALS, or amyotrophic lateral sclerosis, a progressive disease that is slowly breaking down his nerve cells and destroying his muscle control. The condition is incurable, and Immendorff was willing to try nearly anything to alleviate his symptoms. "I see myself as part of an experiment," he told the German newspapers.

The artist's destination was the clinic of Hongyun Huang, a neurosurgeon at Chaoyang Hospital in Beijing. It was Huang who developed the treatment — and it was Huang who injected two million cells from an aborted fetus into Immendorff's brain.

Since 2001, Huang has used cells taken from nasal tissue in fetuses to treat some 600 patients with a number of neurological problems, from spinal injuries to degenerative disorders such as multiple sclerosis and Parkinson's disease. He claims a high success rate, but says he uses the treatment only as a last resort. "Many diseases have no cure," he says, "and we need to try new methods."

Many of Huang's patients are effusive in their praise for the unconventional therapy. But it remains anathema outside China — his critics argue that his results are supported by little more than anecdotal evidence.

This divide presents something of a dilemma. Medical scientists want to see new treatments put through rigorously controlled studies and are suspicious of claims not backed by scientific evidence. A strong patient following is not enough to demonstrate the efficacy of a novel approach, particularly given that desperate patients are eager to try anything that may help. On the other hand, a physician who believes a treatment is working might resist performing placebo controls as these seem unethical. This can leave patients caught in

the middle — unsure of the science, they will often seek the treatment whether or not the research has been done.

Among his patients, Huang has treated about 400 with spinal cord injuries and 100 with ALS. Some who have received the therapy claim that they have improved hand motion and feeling in their fingers. Others report better bowel control and greater ease of breathing. Huang admits that not all of the patients who receive the US\$20,000 treatment respond, but he claims that most do. And he says there is no evidence of serious side effects. As a result, his reputation among patients has spread. Some 3,000 Chinese and 1,000 foreigners are now lined up for the fetal-cell injections.

On a mission

Distracted in an office bustling with assistants, Huang says he is confused over why the Western academic world won't recognize him. He has published nine papers in Chinese journals, including one written in English¹. But despite his claims of clinical success, Huang says that his latest manuscript has been rejected by several of the top international journals, including *Nature Medicine*, *Science*, *The Lancet*, *The New England Journal of Medicine* and the *BMJ*. Huang says his results should be seen as a revolution in clinical treatment, and that rejection is pure discrimination — a refusal to take a serious look at a new method developed by a researcher in China. "If this came from Eng-



Hongyun Huang (left) claims that his technique of injecting fetal cells into patients can improve the quality of life for people with spinal injuries.

land or the United States, it would have already been published," he says.

Many observers disagree with that assessment. They say that the real problem with Huang's work, which he has presented at several international meetings, is the difficulty of measuring his results. For the most part, they are anecdotal. Huang uses videotapes of individual patients and relies on their own assessment of their progress. Such self-assessment is not always consistent. German newspapers reported that Immendorff, for example, felt he could move his hands better after the surgery. But later he said that the treatment hadn't really had any effect.

Unlike proposals to use stem cells to treat conditions such as ALS, Huang's cell injections are not intended to replace damaged nerve tissue. Instead he uses olfactory ensheathing cells (OECs) to promote healing through the production of proteins called growth factors that stimulate cells to grow and divide. OECs are believed to be responsible for the olfactory system's unique and still mysterious ability to sprout new connections to the central nervous system throughout a person's lifetime².

Studies in animals suggest that OEC transplants can help to repair damaged nerves. This was most dramatically illustrated in 2000, when Almudena Ramón-Cueto, a neuroscientist at the Autonomous University of Madrid, used the cells to give paralysed rats the ability to use their legs³. Geoffrey Raisman, a neurobiologist at London's National Institute for Medical Research, has seen similar results. Raisman has been working with OECs since the 1980s, and has found that the cells can significantly improve breathing difficulties in rats with spinal injuries^{4,5} — in humans, breathing problems can chain patients with spinal damage to a ventilator.

All on tape

In his office, Huang plays videos of patients taken before and after the OEC transplant. Before, one ALS patient moves forward in a jerky motion and needs a crutch. His speech is limited to unintelligible mumbling. A month after surgery, the same patient gives a clearly enunciated introduction and, after a struggle to stand, moves forward without a crutch and more smoothly than before. Another who couldn't hold a cup for ten years is shown, post-surgery, signing his name with a pen.

More quantitatively, Huang scores some patients on tests designed by the American Spinal Injury Association (ASIA) and the International Medical Society of Paraplegia. Using the ASIA test, which measures the patient's ability to move, and to feel a light touch and pin pricks, Huang has reported some improvement after the transplant in patients with spinal-cord injury¹.

Since Huang began performing the operations, enthusiastic reports from patients have been broadcast in news reports and Internet discussion groups. Among his champions is Donald Debolt from east central Illinois, who was paralysed in an accident in 1994. He says that he is still improving following his surgery in May 2004. "I was recently able to fasten a button for the first time in 11 years," Debolt says. Increased muscle development in his arms, back and abdomen are a great help when his wife struggles to get him in and out of his wheelchair, he adds.

Laura Jackson, who was paralysed below the neck in a cheerleading accident in May 2003, couldn't breathe on her own for more than five minutes before the surgery. Now she can stay off the ventilator for two hours at a time, and she has regained some trunk movement and sensation in her fingertips. "Was all of it from the therapy? No. Was most of it? Yes," says Daryl, Laura's father.

But not all patients are satisfied. Francis Catena, a major in the Belgian army until a fall broke his spine, says that despite modest improvements in movement and sensation, his therapy was not worth the hassle or the money. Reached at his home in Thailand, Catena complained of unsanitary hospital conditions in Beijing. He adds that his own doctor raised



Chaoyang Hospital in Beijing, where Hongyun Huang carries out his fetal-cell transplants.

questions about whether the surgery had been done properly. "I've seen amazing recoveries from Huang's procedure with my own eyes," Catena says, "but I expected more."

Patient testimonials and video tapes hold little sway with Huang's research peers. According to Raisman, the fundamental problem with Huang's studies is the lack of proper controls or any independent analysis of the data. "He needs to carry out studies for a significant length of time before and after the surgery with a degree of independent assessment," Raisman says. "His studies are not advancing things."

Hard to measure

Martin Schwab, a neuroscientist at the University of Zurich, agrees that lack of careful quantification of the patients leaves evaluation of the procedure reliant on reflex testing and other subjective factors that might produce a placebo effect. "Videotapes and anecdotal evidence are clearly not acceptable as scientific evidence," he says. More compelling evidence

"There are more animal and mouse studies than the world can take. We're ready for the real thing."

— Daryl Jackson

might be possible if Huang used magnetic resonance imaging or electrical recordings of muscle activity to demonstrate changes in neural circuitry, suggests *Nature Medicine's* editor Juan Carlos López. "These would be quite convincing data," he says.

Without such studies, it is difficult to know whether the surgery is worth the risks. Although a number of animal studies using OEC to treat spinal-cord injury have been published, work in animal models of ALS or other neurodegenerative diseases has yet to reach this stage. Such work is normally a prerequisite for gauging the safety and efficacy of a treatment in humans. Furthermore, as the

fate of the cells once they are injected is not well understood, there is cause to worry about side effects, Schwab notes. "They could trigger inflammatory processes which, within the intact central nervous system tissue, would have catastrophic consequences," he says.

Huang counters that he has received approval from his hospital's ethical review board, and he insists that the safety of the procedure has been proven. Although he concedes that there is room for improvement — so far he has not used tissue-type matching to reduce the chance of the cells being rejected, for instance — he argues that it is time for action, not more animal studies. Scientific or not, that argument has a strong appeal to patients. "There are more animal and mouse studies than the world can take. We're ready for the real thing," says Daryl Jackson.

Huang's detractors say that tissue matching, as well as controlled studies, should have been a top priority from the very beginning. They add that this gap between Huang's approach and that of the mainstream biomedical research community will only make it more difficult for him to win general acceptance for his therapy — and to get published.

In the face of the repeated rejections, Huang says he is going to give up trying to convince a Western scientific community that, he is convinced, is prejudiced against him. "It's their loss. If they believed my results, it could dramatically change clinical practice," he says. And the patients are lining up whether or not his work is published.

David Cyranoski is Nature's Asian-Pacific correspondent. Additional reporting by Sonja Schubert and Quirin Schiermeier in Munich.

1. Huang, H. et al. *Chin. Med. J.* **116**, 1488–1491 (2003).
2. Ramón-Cueto, A. & Valverde, F. *Glia* **14**, 163–173 (1995).
3. Ramón-Cueto, A., Cordero, M. I., Santos-Benito, F. F. & Avila, J. *Neuron* **25**, 425–435 (2000).
4. Li, Y., Field, P. M. & Raisman, G. *Science* **277**, 2000–2002 (1997).
5. Li, Y., Decherchi, P. & Raisman, G. *J. Neurosci.* **23**, 727–731 (2003).

BUSINESS

Innovation endgame

The commercial practices of some universities are quietly being transformed by an international chess grandmaster, as Jim Giles reports.

David Norwood, a former captain of the English chess team, has a strong pedigree in technology transfer. Brokers recall Norwood's name from 2000, at the height of the last stock-market boom, when he sold off a technology consultancy firm for a reported £34 million (US\$60 million) — just seven months after he had founded it.

Now Norwood is focusing on the academic world. His company IP2IPO has struck deals with British universities that want help in getting their spin-off companies up and running.

In the United States, some academics meet venture capitalists every other day. That's not the case in Britain, however, and university technology-transfer offices face a hard slog in trying to raise funds to back new spin-off companies. IP2IPO offers to short-circuit the process by signing broad, long-term agreements with universities which allow it to take the lead in exploiting their research.

Right chemistry

Norwood's largest and best-known deal was struck five years ago with the University of Oxford, which had then managed to raise only two-thirds of the £60 million needed to build a new chemistry building. The newly founded IP2IPO offered it the missing £20 million in return for half of the university's stake in all chemistry spin-offs until 2015.

Some financial observers were sceptical — particularly a short time afterwards, when the dot-com bubble burst — about the potential of the deal to unearth discoveries worth investing in. And some Oxford chemists feared that they would be pressured to move into more applied work.

Norwood did not respond to enquiries from *Nature* about how well the agreement has worked. Although analysts, and some researchers, remain cautious, the signs are that both sides have done well so far. "There were concerns when the deal was finalized, but those fears were not realized," says Colin Bain, a surface chemist who left the University of Oxford



Winning move: a company founded by David Norwood (right) is bolstering technology transfer at British universities.

last month for the University of Durham. "There was no pressure from IP2IPO to change the direction of research." And the investment company has already helped Oxford academics to launch eight new companies, including the drug-discovery outfit VASTox, which raised £15 million when it floated last October.

VASTox specializes in screening large numbers of different molecules against zebrafish and fruitfly embryos. Since many genetic sequences are conserved across species, pharmaceutical firms can use the information to guide drug-development work. But IP2IPO's portfolio is not restricted to the life sciences: last year, for example, it bought one-third of Perpetuum, a University of Southampton spin-off. Perpetuum develops devices that generate electricity from vibrations, and these can be used in sensors to monitor machinery.

The deal with the University of Oxford got IP2IPO a lot of attention, and other agreements with King's College London, and the universities of Southampton and York soon followed. In these cases, however, the company has agreed to create a fund that will invest in future spin-offs, rather than giving cash up-front to the university. IP2IPO also bought Techtran, a company set up to operate in association with the University of Leeds. Techtran provides technology-transfer services to academics at the university in exchange for a stake in companies that they create.

The universities that have teamed up with

IP2IPO think it brings them a useful package: valuable seed-funding as well as business advice that should help the companies they set up prepare to go public.

But some of the larger UK universities are taking a different path. Imperial College London has created a company of its own, Imperial College (IC) Innovations, to do roughly what IP2IPO does. This April, the university raised £20 million by selling a 29% stake in IC Innovations to outside investors. Susan Searle, its chief executive, says that her firm provides an integrated service to academics, from identifying market potential to securing venture capital. If all goes well, she aims to float IC Innovations within three years.

Dare to dream

And the University of Cambridge — perhaps the most commercially successful research university in the United Kingdom — has no immediate plans to team up with IP2IPO. "At Cambridge it would be a non-starter," says Ian Leslie, the university's vice-chancellor for research. He thinks that researchers would be uncomfortable with deals that promised stakes to any single outside company.

Cambridge academics like to keep control over the commercialization of their work, he says: some prefer to put their intellectual property into the public domain. But Leslie acknowledges that raising seed-funding is a problem and adds that he might consider a deal like that reached between IP2IPO and the University of Southampton at some stage in the future.

Graham Richards, chair of IP2IPO and head of the University of Oxford's chemistry department, says his company is currently considering several such potential deals and should sign at least one before the end of the year. "We want to turn IP2IPO into a £1-billion company," he says. Interim results released last month show that it earned £3.5 million in the first six months of this year, and now has stakes in 33 university spin-offs. So far, says Richards, "it's been better than we dared dream".

With stakes in a healthy roster of companies, Norwood and Richards' hopes could yet be realized. But others who know the market sound a note of caution. One UK-based venture capitalist, who asked not to be named, claims that British universities are creating more companies because of government pressure to do so, not because there has been a boom in profitable ideas. He estimates that less than 5% of all university spin-offs make it to flotation, with many floundering before even attracting much venture capital. IP2IPO may have shaken up the UK technology-transfer scene, but it remains to be seen whether Britain's universities can create enough wealth to make a success of Norwood's game plan.

H. MILLIGAN

Media should campaign on the basis of facts

SIR — Your Editorial on how the scientific community should respond to public controversies (“Responding to uncertainty” *Nature* 437, 1; 2005) suggested that researchers should attack, on a “scientific basis”, misleading reports that appear in sections of the media. I would like to make some further points about the responsibilities of the media.

Certainly the media should be free to report the opinions of maverick researchers, no matter how unrepresentative these may be of the rest of the scientific community. After all, mavericks are occasionally right. But it is not in the public interest for the media to present these views in a way that creates a misleading impression about the amount of support they have among other scientists.

“The scientific community should make every effort to ensure that the media and public know where the weight of opinion lies.”

— Robert May

As you point out, journalists often strive to achieve a balance by reporting one view and then presenting a diametrically opposed counter-view. When presented in the same way time after time, this can make the research community seem to be evenly divided, even if there are a thousand on one side and one on the other.

As researchers at Cardiff School of Journalism have shown, much of the public gained the wrong impression about how few scientists believed that the triple vaccine for measles, mumps and rubella (MMR), mentioned in your Editorial, was linked to autism and bowel disorders in children, partly because of campaigns in the media. Even more difficulties are created if there are many differing viewpoints.

But the real problems arise when parts of the media decide to campaign on an issue. To take some UK examples, such campaigns may be openly declared, as with attempts by *The Sunday Times* in the 1990s to convince its readers that HIV was not the likely cause of AIDS, and by *The Independent on Sunday* at present to promote the view that genetically modified (GM) foods pose an inherent danger to human health and to the environment. Or they may be undeclared: *The Daily Mail*, for example, seems to be running a campaign to deny the existence of a link between greenhouse-gas emissions and climate change.

In any country, such campaigns can mislead the public about where the weight of scientific opinion lies.

The scientific community should

undoubtedly make every effort to ensure that the public and media know where the weight of scientific opinion lies on issues such as the MMR vaccine, GM foods, HIV and climate change. But surely the media also have a responsibility to find out and convey that information as well?

Robert May

The Royal Society, 6–9 Carlton House Terrace, London SW1Y 5AG, UK

No evidence for Croatian race claims

SIR — I reject the accusation of racism implicit in your News story “Race claims spark fury over Croatia’s school curriculum” (*Nature* 437, 463; 2005). Claims that I am getting teachers to “promote the view that Croats are only distantly related to other Slavic populations such as the Serbs” and that I believe “Croats are more similar to Finns than other Slavs” come from newspapers that misinterpreted my words. Such claims reflect neither my published research nor my subsequent statements and actions on this issue.

The Institute of Education and other authorities, not the minister of science and education, create details of the school curriculum in Croatia. It is up to them to determine which examples of applied science should be offered to Croatian students. Had I, as minister of science and education, either approved or disapproved of the use of data from my scientific work, I could rightly have been accused of meddling. I did not, and any accusations are therefore unjustifiable.

I am deeply disappointed that *Nature* invited my colleagues and collaborators to comment on statements I never made. Your assertion that “scientists inside Croatia are cautious about engaging in public criticism” and its implication about the democratic climate in my country are neither true nor supported by any evidence.

Dragan Primorac

Ministry of Science, Education and Sports, Trg hrvatskih velikana 6, 10000 Zagreb, Republic of Croatia

Katrina: don’t blame the Bush administration

SIR — Your Editorial “Small-minded government” (*Nature* 437, 169; 2005) accuses the US government of failing to “protect its most vulnerable citizens” in the aftermath of Hurricane Katrina. The United States has a federal government, which means that the responsibility for citizen protection, and for first response after the hurricane, lies with the local state and city governments of Louisiana

and New Orleans. We are not a dictatorship in which the national government can override local authorities.

Furthermore, your comment that US leaders may be forced to confront the poverty that was a contributing factor to the crisis after Katrina is unfair. You assume that everyone, given the opportunity, would want to compete in the marketplace to gain a level of economic success measured by an abundance of material goods. There is a limit to what the national government can do — previous efforts have rarely been very successful. This country has drug, and similar, problems that primarily affect the poor, meaning they make poor choices that limit their own economic well-being. But President Bush and his administration have already done something to benefit all the people in this country, including the poor — he has spearheaded the reduction in taxes to stimulate the economy.

The opinion expressed in your Editorial shows a lack of understanding, a judgemental attitude, and a disregard for the elementary rule of good scientific inquiry. It leaves me with a strongly held belief that foreigners generally know very little about this country, and that what they do know is garnered primarily from leftist, elitist news media and fictional movies.

Stephen F. Larner

Department of Neuroscience, Center for Traumatic Brain Injury Studies, McKnight Brain Institute of the University of Florida, Box 100244, Gainesville, Florida 32610, USA

Katrina revealed need for reform. Let’s not forget it

SIR — As an American, I am stunned by *Nature*’schutzpah in editorializing on the performance of the US government in response to Hurricane Katrina (“Small-minded government” *Nature* 437, 169; 2005). And as a patriotic citizen, I am gratified by the boldness of your statement and the insight of your critique.

I hope that our domestic technical publications and professional societies will join *Nature* in calling clearly for reform. In particular, the politically motivated abolition in 1995 of the Congressional Office of Technology Assessment was a costly mistake. The present disaster shows that ignoring warnings by technical professionals does not make problems go away.

Alison Chaiken

47 Esparito Avenue, Fremont, California 94539, USA

Nature’s Editorials are written by editorial staff. As in this case, writers include US citizens and others who have lived in the United States for years — Editor, *Nature*.

BOOKS & ARTS

A secular religion

Should evolutionism be viewed as a modified descendant of Christianity?

The Evolution–Creation Struggle

Michael Ruse

Harvard University Press: 2005. 336 pp.
\$25.95, £16.95**John Hedley Brooke**

Few disputes have generated as much emotion, bitterness and incomprehension as the enduring conflict between darwinians and their creationist opponents. Those conversant with classical Christian theology know that the doctrine of Creation speaks of the ultimate dependence of everything on a transcendent power: it doesn't need supernatural conjuring tricks to account for each new species. Even Charles Darwin's 'bulldog', T. H. Huxley, insisted that the theory of evolution had no more implications for theism than had the first book of Euclid. Why then, and in North America especially, has there been such a highly polarized, emotive and undiminished debate?

In *The Evolution–Creation Struggle*, Michael Ruse tries to explain this puzzling cultural phenomenon. He is well known as a committed darwinian philosopher, experienced in gutting claims that creationism and 'intelligent design' can be a form of science. His aim in this book, however, is not to attack but to understand. For that he wisely turns to history — specifically to the history of evolutionary theory itself and the cultural contexts in which it was forged, refined and publicized.

The purpose of Ruse's admittedly streamlined history is to identify two divergent responses to a crisis in Christianity arising from Enlightenment critiques. One response was a belief system in which a high value was placed on social and intellectual progress, into which ideas of biological progress (and eventually a science of evolution) would comfortably fit. The other response was a mutation of Christianity itself, epitomized by the evangelical spirit of Methodism, a defensive attitude to the authority of the Bible, and a millenarian vision in which, after testing man's devotion, God would allow the return of Christ for a 1,000-year rule of a perfected human society.

Ruse's argument is that these antithetical responses graduated into the two competing world-views that lie at the heart of the contemporary conflict. His thesis leads to a radical conclusion. Although we are used to speaking of a conflict between science and religion, to do so misses the point: it is rather a conflict

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

No need for conflict: T. H. Huxley believed that evolutionary theory has no implications for belief in God.

between religion and religion, he claims. There is a sense in which it is an intra-family feud, and this explains its bitterness.

Superficially this may sound paradoxical, if not perverse. Surely scientific theories of evolution cannot be paraded as examples of religious belief? Of course not. But Ruse has in mind a distinction between evolution as a fact, evolution as a theory that offers mechanisms for evolutionary change, and 'evolutionism' — a metaphysical, naturalistic world-view imbued with values as well as a strictly scientific narrative. It is evolutionism that has repeatedly functioned as a secular religion, offering seductive images of progress and translating naturalistic methods of enquiry into doctrinaire assertions about what can and cannot be believed about the meaning of human existence.

Ruse asserts that for many evolutionary

biologists, "evolution was their profession...evolutionism their obsession". From the earliest prominent evolutionists (Erasmus Darwin, Jean-Baptiste Lamarck and Robert Chambers) to latter-day darwinians such as Richard Dawkins, proponents of biological evolution have tended to be deists or free-thinkers who have self-consciously rejected Christianity, only to replace it with a substitute system that presumes to answer the same basic questions.

As justification for treating evolutionism as a religion, Ruse observes that it supplies a story about origins; it reaffirms a unique role for humans in shaping the future; it has not uncommonly made moral prescriptions (some, such as eugenics, now blacklisted); it has opposed other religious systems; and, with recurrent insistence on progress, it has its own view of how the world might end. Strikingly, the language used by champions of an evolutionary world-view underlines its religious character. Ruse quotes Dawkins: "All the great religions have a place for awe, for ecstatic transport at the wonder and beauty of creation.

And it's exactly this feeling of spine-shivering, breath-catching awe — almost worship...that modern science can provide." Ruse makes no secret of his admiration for E. O. Wilson, whose call to repentance on the subject of biological diversity reminds him of an old-time preacher.

Whether a secular world-view should properly be described as 'religious' is ultimately a matter of definition. In reflective moments, Ruse opts for the qualified 'quasi-religious'. Ultimately, his justification for such labels stems from an insight that I first encountered in E. L. Tuveson's study *Millennium and Utopia* (University of California Press, 1949). This is that the modern idea of progress arose in seventeenth-century Europe through a secularization of millenarian theology. Biblical texts were reinterpreted to suggest that through human effort, including scientific and technological innovation, the Earth could be

brought to a state of paradise appropriate to Christ's return. To some this will sound quaint. But the fact remains that a science-based utopia, with its supreme confidence in the technological fix, had at least some roots in a newly conceived religious imperative: to hasten the millennium.

It is therefore possible to argue, as Ruse does, that the struggle between evolution and creation is a contest between rival versions of millenarian theology. The world-view and conservative moral values of the creationists tend to be informed by a theology in which God alone can instigate a more perfect society through human redemption. The world-view of the popularizers of evolutionary biology tends to be informed by the legacy of an alternative reading in which humans had to take responsibility for shaping the future. Not for nothing did Julian Huxley describe his evolutionary humanism as a "religion without revelation".

Ruse knows that not all of Darwin's disciples can be shoehorned into his scheme. Indeed, he takes trouble to discuss exceptions. The late Stephen J. Gould is an obvious one, given his aversion to progressive readings of the fossil record and his equally adamant line in *Rocks of Ages* (Ballantine, 1999) that the respective magisteria of science and religion must not be allowed to overlap. For Ruse, Gould is the exception that proves the rule — not least because, on closer analysis, his reduction of religious provenance to questions of morality places him squarely in the secularizing tradition. And one might add that a close analysis of Gould's controversy with Simon Conway Morris shows that his well-known take on the fossils of the Burgess shale was, by his own admission, informed by his own social values.

This book is aimed at a general audience. Ruse's style is chatty and informal, sometimes incongruously so. His historical treatment of Christianity will be too sketchy for fastidious scholars. But let us not underestimate the importance of his message. His appeal is to all who love science; his exhortation is that we "must do more than simply restate our positions or criticize the opposition". A prerequisite of progress in this cultural struggle is that we should recognize the metaphysical assumptions underlying dogmatic forms of scientific naturalism, and be willing to investigate the concerns that motivate criticism.

Ruse has done his best to reveal both. He declines to be a prophet of doom, but precisely because creationism is linked in America with perceived moral and political threats impinging on society, he cannot see it vanishing any time soon. Insofar as he has a remedy, it is that Christian and secular evolutionists should not waste time sniping at each other but collaborate instead. In this regard he would be pleased to learn that Richard Harries, the bishop of Oxford, has joined the evolutionist Richard Dawkins in alerting the British public to defects in creationist rhetoric (unlike an

unhappy predecessor, Samuel Wilberforce, who in 1860 tried to outwit Huxley). In a joint article published in *The Times* on 10 June 2005, the two Richards gave complementary reasons

why the promotion of creationism in schools should be resisted. ■

John Hedley Brooke is at Harris Manchester College, University of Oxford, Oxford OX13TD, UK.

The rise of the computer

Electronic Brains: Stories from the Dawn of the Computer Age

by Mike Hally

Granta: 2005. 224 pp. £15.99

Published in the United States as *In a Fraction of a Second* by Joseph Henry Press (\$27.95)

Anthony Ralston

Writing a history of events that took place 50 years ago has the advantages that almost all the documentation is available, and that some of the protagonists are alive to be interviewed. Mike Hally has done this for his book about the development of computers in the years just after the Second World War. But the danger of relying too heavily on such interviews is that important topics can be missed if there are no surviving protagonists.

Electronic Brains grew out of four 15-minute programmes that Hally wrote and produced for BBC radio in 2001. As he accumulated far too much material for one hour of broadcasting, this volume contains much more information than the radio series.

When writing about computers, the question arises as to what you mean by 'computer'. Most people today would agree that this word implies a machine that is general purpose (it can be programmed to solve essentially any solvable problem), electronic, has a stored program that is held and executed in the machine, and digital, being based on discrete rather than continuous technology. Even so, it is perfectly reasonable to discuss, as Hally does, the pre-history of modern computers, when they were often special-purpose, sometimes mechanical, and seldom stored-program. It is nevertheless surprising to find a chapter devoted to MONIAC (Monetary National Income Automatic Computer), an analogue, special-purpose, hydraulic, externally programmed device built in England for economic modelling. But the chapter is fascinating, and I hurried over to the Science Museum in London to view its MONIAC exhibit.

The chapters on some very early computers, UNIVAC (Universal Automatic Computer, Eckert-Mauchly Computer Corporation, USA, 1951), EDSAC (Electronic Delay Storage Automatic Calculator, University of Cambridge,

1949), and LEO (Lyons Electronic Office, an adaptation of EDSAC for business applications, 1951), are competent expositions about three of the first computers to satisfy the definition in the previous paragraph. This material will be familiar to most computer scientists but not to laypeople, who will find much of interest about the development of these computers and the people who developed them.

The chapter on the RAND 409, a computer built by Remington Rand of Rowayton, Connecticut, is another matter. I admit to prior ignorance of this computer, but the startling thing about this fascinating chapter is that neither here nor in the later chapter on IBM is there any mention of the IBM Card Programmed Calculator (CPC). The CPC was contemporary with, and similar to, the 409, sold better, and was much more influential in the later development of computers. The CPC rates a full page in the 1996 book *Computer: A History of the Information Machine* by Martin Campbell-Kelly and William Aspray (Basic Books), whereas the 409 is not mentioned at all.

Hally's book also contains chapters on the pre-history of computers in the 1930s and 1940s, as well as on early computers in the Soviet Union and Australia.

The final chapter, on the rise and dominance of IBM, is most noteworthy for the remarkable claim that IBM "doesn't figure much" in the early history of computers, "at least until the mid-1960s". Actually, as Hally later makes clear, by the early 1950s IBM was already abreast of its only real competitor, Remington Rand, which bought the Eckert-Mauchly firm in 1950. By the late 1950s, IBM was already in a class by itself.

Hally is quite accurate about hardware but, perhaps because he is an engineer, less so about software. One blooper: the American journal *Mathematical Tables and Other Aids to Computation*, published by the National Research Council in the 1940s and 1950s, is referred to as a "Russian journal". Still, this book contains enough fascinating material to be of interest to both computer experts and non-experts. ■

Anthony Ralston is professor emeritus at the State University of New York at Buffalo. His mailing address is Flat 4, 58 Prince Consort Road, London SW7 2BE, UK.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Big deal: UNIVAC was among the first stored-program computers.

BETTMANN/CORBIS

EXHIBITION

Lighting up the background

Intermittent flashes of light were seen emanating at night from the physics department at Queen Mary, University of London, for six weeks last year. They were triggered by background radiation: subatomic particles detected by a hundred Geiger counters connected to photography flash-bulbs.

The installation *What the Eye Can't See the Heart Can't Grieve For* — devised by artist Matthew Tickle, in collaboration with physicist Fay Dowker, who seeks to interpret quantum physics — contradicts the adage provided by its title. It reminds us that the invisible can be dangerous.

The experience is now captured on DVD, where some images, such as the one shown here, evoke the surreal façades painted by René Magritte. An accompanying booklet remarks that "the work exists only as a series of moments of perception, as imprints on our retinas. The moment of looking is the point at which the art exists."

C.M.

www.mattsgallery.org



MATTHEW TICKLE/MATT'S GALLERY, LONDON

A philosopher's vision

Action in Perception

by Alva Noë

MIT Press: 2005. 392 pp. \$38, £24.95

Richard L. Gregory

Philosophers contribute significantly to the brain sciences by clarifying terminology and concepts, occasionally issuing radical challenges and stimulating suggestions. It might be said for the whole of science that conceptual significance is as important as statistical significance, and philosophy contributes most in areas where generally accepted paradigms are lacking, such as cognitive brain science. The mystery of consciousness makes scientists listen to even the wildest ideas and most extreme challenges of philosophers. These are most effective when the philosopher has a grasp of the physiology and key experimental data.

Action in Perception by the philosopher Alva Noë is a short and clearly written book that abounds in such challenges. Like books by Daniel C. Dennett, the pioneer in this area, it discusses both phenomena and experiments. No doubt brain scientists would have liked the brain to be central to the book, but although it is mentioned occasionally, 'brain' does not even qualify for a place in the index.

Noë objects to theories of vision as 'snapshots', stressing that it is built up from many fixations of the moving eyes — rather like discovering the form of objects by exploring them by touch. Touching bits of a bottle with the eyes shut is sufficient to experience the whole bottle. If you look at the front surface of a tomato, you see the whole tomato as an object, even appreciating what is inside. Noë makes a useful distinction between 'seeing' and 'presence'.

This is really a battle with straw men, however, as comparing seeing with touching goes

back to George Berkeley in the eighteenth century. The comparison was also made in the middle of the nineteenth century by Hermann von Helmholtz, the founder of the science of visual perception. Von Helmholtz stressed the importance of eye movements (which until recently were hard to record, but were not ignored), and the parallels between active touch and vision, in an eloquent account of the importance of explorative touching in helping children learn to see and understand. For followers of von Helmholtz, it is not only the extra information provided by touch that is important for seeing — knowledge of many other non-optical features, including those sensed by taste, smell and sound, contribute to making the tenuous images in the eyes useful.

Berkeley's concept of a blind man exploring spatial relations by tapping his stick is developed differently by Noë in his 'enactive theory'. This ambitious theory tries to make seeing even more closely related to the object world than touching, despite the intervening retinal image and the physiological complexity of the visual channels and brain processing. It claims that seeing colours is similar to feeling objects: for Noë, colours exist objectively in objects, so there would be colours in the world even if there were no eyes or brains. This was denied by Newton and John Locke in the seventeenth century, and by current brain science, which posits that colours are created in brain regions, although how this happens is still a mystery, as consciousness is not understood.

Philosophy and science do battle over phenomena such as not seeing a blank hole off the centre of this page, where the image falls on the 'blind spot' of the retina. Although there are no retinal signals from this large region, we do not see blackness — or nothing — but a

complete page. Is the missing region actively filled in, by extrapolation from its surroundings? Or is it, as Dennett and Noë believe, passively ignored (like a boring person at a party), as it never provides useful information? This point is well worth considering, as it is easy to bark up the wrong tree, or (if you will) to bark up trees that do not exist; extensive barking may suggest the presence of non-existent trees, leading to the creation of mythologies. There are experiments that support mechanisms for active filling-in, some published in *Nature*. But the debate over active filling-in versus passive ignoring is a controversial issue for many perceptual phenomena, with very different implications, some of them clinically important.

Philosophers' accounts of consciousness have special interest as experiments are few and far between and are extremely hard to interpret. Perhaps the most interesting discussions in this book are centred on the radical scepticism known as the 'grand illusion', initiated by Dennett and recently discussed by others including Susan Blackmore. The challenging claim is that we are misled by our consciousness into thinking of visual experience as continuous and rich, whereas it is really discontinuous and sketchy. This view is apparently supported by striking phenomena such as change blindness and inattention blindness, which underlie much of conjuring.

Noë's balanced, well-considered account of this hot topic makes the claim — contrary to most brain scientists — that the brain does not provide an internal representation of the world. Surely Noë is right to ask how much is represented by the brain, and how much of what we seem to see is illusion? No doubt this is an empirical question, but philosophy serves a useful purpose in asking it in ways that may be answered experimentally. So philosophy and science meet, to mutual benefit.

Richard L. Gregory is in the Department of Experimental Psychology, University of Bristol, Bristol BS8 1TN, UK.

A quantum recipe for life

Sixty years on, Erwin Schrödinger's prediction that quantum mechanics would solve the riddle of how life started has not been fulfilled. But the appeal of using quantum theory to solve the mystery persists.

Paul Davies

One of the most influential physics books of the twentieth century was actually about biology. In a series of lectures, Erwin Schrödinger described how he believed that quantum mechanics, or some variant of it, would soon solve the riddle of life. These lectures were published in 1944 under the title *What is life?* and are credited by some as ushering in the age of molecular biology.

In the nineteenth century, many scientists thought they knew the answer to Schrödinger's rhetorical question. Life, they maintained, was some sort of magic matter. The continued use of the term 'organic chemistry' is a hangover from that era. The belief that there is a chemical recipe for life led to the hope that, if only we knew what it was, we could mix up the right stuff in a test tube and make life in the lab.

Most research on biogenesis has followed that tradition, by assuming that chemistry was a bridge — and a long one at that — linking matter with life. Elucidating this chemical pathway has been a tantalizing goal, spurred on by the famous Miller-Urey experiment of 1952, in which amino acids were made by sparking electricity through a mixture of water and common gases. But the concept has turned out to be something of a blind alley, and further progress with prebiotic chemical synthesis has been frustratingly slow. The origin of life remains one of the great outstanding mysteries of science.

To take up Schrödinger's suggestion, a radical solution to the problem, 'What is life?' could be that quantum mechanics enabled life to emerge directly from the atomic world, without the need for complex intermediate chemistry. Life must have a chemical basis: organic molecules provide the hardware for biology. But what about the software?

When Schrödinger asked, 'What is life?' he could already glimpse the central significance of the cell's information storage and replication processes, even though the role of DNA and the genetic code was yet to be discovered. Today, the cell is regarded not as magic matter but as a computer — an information-processing and replicating system of astonishing precision.

When life is viewed in terms of information processing, the problem takes on a different complexion. Biologists have always regarded reproduction — one of the defining characteristics of life — as being about replicating structures, whether they be DNA molecules or entire cells. But to get life started all you need is to replicate information.

Information can be processed at the quantum level orders of magnitude more rapidly than it can be processed classically, which is why the race is on to build a quantum computer. Furthermore, quantum systems can make use of phenomena such as superposition, entanglement and tunnelling to enhance their performance.



Finding the atomic Adam: quantum mechanics may have allowed life to emerge without the need for complex intermediate chemistry.

A quantum replicator need not be an atomic system that clones itself. Indeed, there is a quantum no-cloning theorem that forbids the replication of wavefunctions. Rather, the information content of an atomic system must be copied more or less intact — not necessarily in one step, but maybe after a sequence of interactions. This information might well be in binary form, making use of the spin orientation of an electron or atom for example. Quantum mechanics thus provides an automatic discretization of genetic information.

What is this atomic Adam, this quantum replicator that begets life? I confess I haven't a clue about the best environment in which to find such a thing, although I know it would not be in a traditional primordial-soup setting. It might even be a frigid location such as an interstellar grain. Wherever it was, once a population of information replicators became established, quantum uncertainty provided an

inbuilt mechanism for variation. Throw in a selection mechanism and the great darwinian game could begin.

How, then, did organic life arise? Information can readily be passed from one medium to another. At some stage, quantum life could have co-opted large organic molecules for back-up memory. Eventually the organic stuff would literally have taken on a life of its own. The loss in processing speed would have been offset against the greater complexity, versatility and stability of organic molecules, which in turn would have enabled organic life to invade many environments.

Something is missing from the account so far — complexity. Replicating a single bit of information is one thing; generating and replicating long concatenations of bits is quite another. How complexity emerges in quantum systems is a subject still in its infancy, but the principles involved could be illuminated by applying algorithmic complexity theory to quantum information theory.

When Schrödinger published his book, quantum physicists were flushed with the success of explaining the nature of matter. Life is, after all, just a state of matter, albeit a weird one. Sixty years on, Schrödinger's expectation has not been fulfilled.

Molecular biologists are content with ball-and-stick models based on classical concepts. But so long as they cling to that, the origin of life will remain mysterious.

Even if we can't reconstruct the precise details of life's emergence, knowing the general principles would be a huge advance. Proving a quantum-mechanical theorem that puts a bound on the probability that such-and-such a system can replicate to a certain accuracy, and evolve to a particular level of complexity, might answer astrobiology's burning question: was the origin of known life a freak accident, or the expected outcome of intrinsically bio-friendly laws of physics? Momentous implications would flow from the answer, as the issue addresses one of the deepest questions of existence: is life a cosmic phenomenon, or are we alone in the vastness of the Universe? ■

Paul Davies is a physicist in the Australian Centre for Astrobiology, Macquarie University, Sydney, and author of *The Origin of Life* (Penguin, 2003).

M. KUIYK/SPL

ESSAY

MICROBIOLOGY

Loading the type III cannon

Bill Blaylock and Olaf Schneewind

Many pathogenic bacteria possess a secretion machine that shoots noxious proteins into host cells. But the ammunition is larger than the bore of the bacterial gun, so how is it fed into the machine?

Bacteria and the organisms they parasitize are engaged in a constant struggle. Each side possesses a veritable arsenal of weapons and defensive countermeasures. One of the most potent weapons systems that bacteria use is the type III secretion system, which injects certain bacterial proteins into adjacent host cells. The proteins selected by the secretion machinery often resemble the host's own proteins, so they can switch off defensive systems and turn the cells into puppets of a bacterial master. In this issue, Akeda and Galán (page 911)¹ give us a first peek into the loading mechanism of the bacterial type III secretion system.

The crystal structures of proteins that are transported by the type III secretion machine have been solved². The diameter of the conduit these proteins must travel has also been ascertained³. A paradox emerges when these data are compared — the proteins are too large to travel through the secretory apparatus. So how do bacteria get the camel to pass through the eye of a needle? A solution suggested by Akeda and Galán implicates InvC, an essential component of the *Salmonella enterica* type III machinery.

InvC belongs to a class of enzymes called AAA ATPases that form a hexameric ring. These enzymes harness the energy released from ATP to unfold proteins and thread them through a channel at the centre of the ring⁴. Akeda and Galán reasoned that if InvC behaved like this at the breech of the type III secretion machine, secreted proteins would be of the correct calibre to pass through. Their careful *in vitro* examination of the SptP protein, which is transported by the *S. enterica* type III system, supports this idea. Before secretion, SptP is bound by a 'chaperone' called SicP₂, which is essential for its recognition by the secretion machinery. The authors found that InvC binds to the SptP–SicP₂ complex; that the consumption of energy by InvC releases SicP₂; and finally, that the SptP that is also liberated is in an unfolded state¹ (Fig. 1).

Akeda and Galán's hypothesis has several

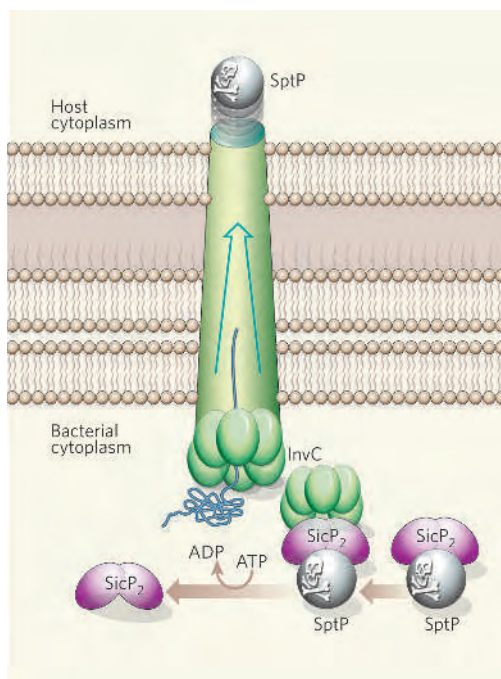


Figure 1 | Calibrating the ammunition. Bacteria shoot toxic proteins (such as SptP) into host cells through a type III secretion machine. In the bacterial cell, SptP is bound by a recognition factor (SicP₂) that docks at the InvC component of the secretory apparatus. InvC consumes energy (ATP) to dissociate SptP from SicP₂ and to unfold SptP — allowing it to fit through the narrow tube connecting the two cells. SicP₂ is released into the bacterial cell, and the toxic SptP is fired into the host cell.

attractive features: InvC is associated with the cell membrane; it could therefore reside at the site where the pipeline into the host cell meets the bacterial cytoplasm, where there is a pool of substrate proteins to be secreted⁵. Moreover, because InvC can indirectly bind to secretion substrates through their cognate chaperones, it performs all the functions that are needed for the priming of proteins for type III travel: recognition, release of chaperones, and unfolding. Other secretion machines may indeed use similar mechanisms for substrate selection by chaperone binding, and these include *Escherichia coli* type III machines as well

as the secretion of proteins that make up flagella (the whip-like 'tails' that propel some bacteria)^{6,7}.

This work may conveniently solve another puzzle: where does the energy for substrate travel come from? If InvC acts as the pumping station at the end of a pipeline, it is conceivable that the energy it invests in pushing substrates in one end is sufficient to shunt previously pumped substrates out of the other.

Given our current understanding of InvC, it is tempting to take the view that secretion substrates dock with InvC through their bound chaperones. Following chaperone displacement, a substrate would then be sucked in through the ring and unwound like a strand of spaghetti. The closely related flagellar assembly and secretion system has an InvC equivalent. In this system, the other components of the secretory machinery that face the interior of the bacterium can recognize not only substrates for secretion but also each other⁸. One of these proteins can even modulate the rate of ATP consumption of the InvC equivalent and another is part of the membrane-embedded secretion machinery^{8,9}. So it seems likely that a complex web of interactions modulating InvC activity and specificity remains to be discovered, thereby providing even greater insight into the mechanisms by which bacteria load their type III weaponry. ■

Bill Blaylock and Olaf Schneewind are in the Department of Microbiology, University of Chicago, 920 East 58th Street, Chicago, Illinois 60637, USA.
e-mail: oschnee@bsd.uchicago.edu

1. Akeda, Y. & Galán, J. E. *Nature* **437**, 911–915 (2005).
2. Stebbins, C. E. & Galán, J. E. *Nature* **414**, 77–81 (2001).
3. Marlovits, T. C. *et al. Science* **306**, 1040–1042 (2004).
4. Sauer, R. T. *et al. Cell* **119**, 9–18 (2004).
5. Pozidis, C. *et al. J. Biol. Chem.* **278**, 25816–25824 (2003).
6. Gauthier, A. & Finlay, B. B. *J. Bacteriol.* **185**, 6747–6755 (2003).
7. Thomas, J. *et al. Proc. Natl Acad. Sci. USA* **101**, 3945–3950 (2004).
8. Minamino, T. & Macnab, R. N. *Mol. Microbiol.* **35**, 1052–1064 (2000).
9. Minamino, T. & Macnab, R. N. *Mol. Microbiol.* **37**, 1494–1503 (2000).

ASTROPHYSICS

Short-burst sources

Luigi Piro

Measurements of the X-ray afterglow of long γ -ray bursts largely clarified the origin of these bright flashes of cosmic radiation. Their shorter-lived siblings are now beginning to divulge their secrets, too.

The distance, energy output and source of the mysterious flashes of high-energy radiation known as short γ -ray bursts have so far resisted all attempts at their elucidation. Observations presented in this issue^{1–4} establish for the first time the cosmological distance of the bursts and provide solid support for the favoured theoretical model for their origin. In this scheme, short γ -ray bursts are the expression of the enormous energy released when two compact objects — remnants of exploded stars in the form of highly dense neutron stars or black holes — merge.

Scientists have been fascinated by γ -ray bursts (GRBs) for more than 30 years. The most glamorous celebrities of the sky, they appear unpredictably, and — after a brief and intense show, during which their brightness overwhelms that of any other celestial source of γ -rays — they disappear. There are two classes of burst, identified according to their duration: long (longer than two seconds) and short⁵. The source of long bursts was revealed in 1997, thanks to the precise and fast localization capabilities of the BeppoSAX satellite put into orbit by the Italian and Dutch space agencies. The result was the discovery of the bursts' fainter, long-lived afterglow emission at X-ray⁶ and optical⁷ frequencies, lower than the original γ -ray emission. The observations linked these events to the explosive collapse of the cores of young, massive stars in sub-luminous, star-forming galaxies at redshifts of 1–2. (The redshift of cosmological radiation is an indicator of a source's distance; a redshift of z indicates that the Universe has expanded by a factor of $1+z$ since the radiation was emitted.)

The long-wavelength afterglow of short GRBs has until now evaded detection. The inference — that their afterglow is simply much dimmer than that of the long GRBs — fits with the merger model favoured to explain short GRBs. This model starts with the supernova explosions of two massive progenitor stars in a binary system, which collapse under their own weight to form extremely dense neutron stars or, in the case of more massive progenitors, black holes. The kick imparted to these compact objects by their supernova explosion sends them into less-dense spaces in their galaxy, where they eventually merge, producing the short γ -ray burst (Fig. 1). The afterglow of a burst is produced when material ejected by the explosion at relativistic speeds interacts with the surrounding medium. Therefore, because the merger occurs in less-dense

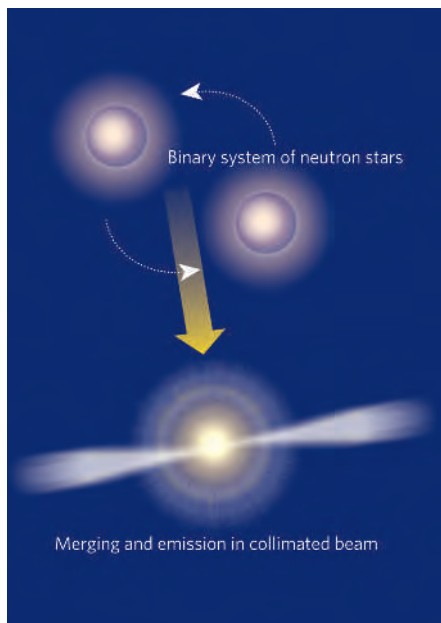


Figure 1 | Bursting onto the scene. The current most favoured model of the formation of short γ -ray bursts (short GRBs) is supported by recent observations^{1–4}. In this model, two extremely dense compact objects — remnants of exploded stars such as neutron stars or black holes — merge, sending out a highly directed emission at γ -ray frequencies.

regions of a galaxy, little or no afterglow emission would be expected from a short GRB.

Nevertheless, when the light curves of tens of short GRBs that had been observed independently by BeppoSAX and NASA's BATSE satellite were summed up, they revealed hard, or high-frequency, X-ray tails lasting several tens of seconds^{8,9}. It was an indication, albeit of limited statistical significance, that some long-lived emission was indeed present, and left open the possibility that the X-ray afterglows of short GRBs had been missed not because of their faintness, but simply because none of the available X-ray cameras was pointing the right way. The launch in November 2004 of Swift, a new NASA satellite carrying a wide-field hard-X-ray imager, was intended to eliminate that possibility.

Elsewhere in this issue, the fast and precise localization of two short GRBs is reported in four papers^{1–4}. Gehrels *et al.* (page 851)¹ report the identification by Swift of a first event (GRB 050509B) on 9 May this year. The satellite was slewed rapidly in the direction of the burst, enabling its narrow-field X-ray telescope to

pinpoint to within 10 arcseconds (1/360th of a degree) a source of faint X-ray emission one minute after the burst. This allowed the authors to assign a likely origin for the burst — a luminous, non-star-forming elliptical galaxy at a redshift of 0.225.

Villasenor *et al.* report (page 855)² the observation of a second short burst, GRB 050709, pinpointed by NASA's HETE2 satellite on 9 July this year. Fox *et al.* (page 845)³ and Hjorth *et al.* (page 859)⁴ found long-lived X-ray and optical afterglows following this burst, and could thus associate it unambiguously with a star-forming galaxy at a redshift of 0.16. This second event would thus seem at first glance to share some properties — a definite afterglow and an origin in a star-forming galaxy — with long GRBs.

Taking the distance of its source into account, however, the luminosity of the X-ray afterglow of GRB 050709 is, in fact, some three orders of magnitude lower than that typical of a long burst. Short GRBs might also be expected to occur in younger galaxies where stars are still forming, as well as in elliptical galaxies (which tend to consist of older stars), as anything from a few million to several billion years can separate the formation of the compact objects and their merger¹⁰. Having said that, two more short bursts localized by Swift in recent months (see ref. 3 and references therein) also seem, like GRB 050509B (ref. 1), to originate in elliptical galaxies — a predominance that seems to indicate that the progenitor stars on average evolved over billions of years.

The energy output of the short GRBs can be easily derived once the distance of the source is established (assuming the emission is isotropic), and is found to be in the range of 10^{48} – 10^{50} erg. Again, this is around 1,000 times lower than that of long GRBs. It is nonetheless large enough to rule out flares from highly magnetized neutron stars, known as soft γ -ray repeaters (SGRs)¹¹, as the explanation for the bursts. It does not, however, preclude the possibility that some short bursts might be the product of an SGR in a nearby external galaxy.

The properties of the long-lived afterglow emission observed for GRB 050709 are consistent with the possibility that the emission arises from low-density surrounding material that is shocked by a collimated (directed) relativistic blast wave — the same process thought to be at work in the long bursts. But the lower energy of the short bursts, together with the fact that they originate closer to Earth in a mixture of galaxy types, suggest that short and long GRBs are in fact separate populations. The observed characteristics^{1–4} of the short GRBs are all consistent with models of the merger of two neutron stars, or of a neutron star with a black hole.

Several questions remain. First, if short bursts are preferentially associated with nearby bright elliptical galaxies, why has no such

galaxy been found previously in locations where short GRBs have been well defined¹²? This may be because the population of progenitors had different properties, or because these were distant events originating in faint galaxies that have yet to be identified. Second, why does the afterglow of GRB 050709 show³ evidence of flaring X-ray activity two weeks after the short burst? This is particularly puzzling because it suggests a long-lived energy injection from the central engine driving the burst.

As occurred with their longer cousins, the discovery of long-lived afterglows of short GRBs sets the stage for detailed studies of these exotic cosmic explosions through their emissions of electromagnetic radiation. In the future, there is the exciting prospect of detecting gravitational waves from these events

using the second generation of laser interferometry detectors, LIGO and VIRGO. ■

Luigi Piro is at the Istituto Astrofisica Spaziale e Fisica Cosmica, INAF, Via Fosso del Cavaliere 100, 00133 Rome, Italy.
e-mail: piro@rm.iasf.cnr.it

1. Gehrels, N. *et al.* *Nature* **437**, 851–854 (2005).
2. Villaseñor, J. S. *et al.* *Nature* **437**, 855–858 (2005).
3. Fox, D. B. *et al.* *Nature* **437**, 845–850 (2005).
4. Hjorth, J. *et al.* *Nature* **437**, 859–861 (2005).
5. Kouveliotou, C. *et al.* *Astrophys. J.* **413**, L101–L104 (1993).
6. Costa, E. *et al.* *Nature* **387**, 783–785 (1997).
7. van Paradijs, J. *et al.* *Nature* **368**, 686–688 (1997).
8. Lazzati, D., Ramirez-Ruiz, E. & Ghisellini, G. *Astron. Astrophys.* **379**, L39–L43 (2001).
9. Montanari, E., Frontera, F., Guidorzi, C. & Rapisarda, M. *Astrophys. J.* **625**, L17–L21 (2005).
10. Fryer, C. L., Woosley, S. E. & Hartmann, D. H. *Astrophys. J.* **526**, 152–177 (1999).
11. Hurley, K. *et al.* *Nature* **434**, 1098–1103 (2005).
12. Hurley, K. *et al.* *Astrophys. J.* **567**, 447–453 (2002).

MICROBIOLOGY

Conspirators in blight

Ian R. Sanders

A fungus and a bacterium have been found in a symbiotic alliance that attacks rice plants. Rice feeds more people than any other crop, but the significance of this finding extends beyond its potential agricultural use.

Rice suffers from a serious disease called seedling blight. The cause was thought to be a toxin released by some species of the fungal group *Rhizopus*. The toxin kills root cells, after which the fungus digests the remains of the dead root. But as Partida-Martinez and Hertweck report on page 884 of this issue¹, that is not the case — they find that the toxin is produced not by the fungus, but by bacteria that live in symbiosis inside it.

This finding may help in controlling seedling blight, no minor consideration given that rice feeds more people in the world than any other crop plant (Fig. 1). Moreover, the toxin — rhizoxin — stops cell division in some lines of human cancer cells. It is under investigation as a potential antitumour agent, so identification of the genes involved in rhizoxin production could also provide lessons for cancer researchers.

An enzyme, a polyketide synthase (PKS), has been implicated in the biosynthesis of rhizoxin, which is associated with only some species, or strains, of *Rhizopus*. But when Partida-Martinez and Hertweck looked for fungal PKS genes in the genome of *Rhizopus* strains known to release the toxin, they could not find them. However, the authors did detect PKS genes in the fungus that were similar to a known class of bacterial PKS genes.

The next and obvious question was whether these genes really exist in bacteria living inside the fungus. Partida-Martinez and Hertweck first amplified and sequenced a particular set

of genes — 16S ribosomal genes, unique to bacteria — from DNA extracted from the fungus (which would include DNA of any bacteria living inside it). They found that the 16S sequences belong to bacteria of the genus *Burkholderia*, a group that occupies a remarkably wide range of ecological niches². *Burkholderia* 16S genes were not found in strains of *Rhizopus* that do not release the toxin.

Using laser microscopy, the authors were

able to observe bacteria living inside toxin-producing *Rhizopus* strains. They then treated these strains with antibiotics and showed, again with laser microscopy, that the fungus no longer contained the bacteria. Bacterial genes were not evident in these bacterium-free strains, and the fungus did not release detectable amounts of rhizoxin.

From these findings, it seemed likely that *Burkholderia* bacteria were producing the toxin that causes rice seedling blight. But Partida-Martinez and Hertweck went further. Bacteria that live inside other organisms are notoriously difficult to culture on artificial media, and many are considered unculturable³. However, Partida-Martinez and Hertweck successfully isolated the bacteria and cultured them without their fungal host. They showed that it was indeed a pure culture of a species of *Burkholderia* and that it produced the toxin in artificial media. When the bacteria were reintroduced into a bacterium-free fungus, the fungus again produced significant amounts of rhizoxin. Although the toxin stops cell division in human cells, for example, it clearly doesn't harm the fungus.

The exciting aspects of this research¹ go beyond the prospects for controlling seedling blight in rice and using rhizoxin to treat cancer: the existence and evolution of such a symbiosis between the fungus and bacterium are in themselves intriguing. Close relatives of *Burkholderia* are well known as symbionts that commonly live inside other fungi, called arbuscular mycorrhizal fungi, which in turn live symbiotically in the roots of most plant species^{4,5}. The role of these bacteria in mycorrhizal fungi has remained elusive because of the difficulty of culturing them. So the new work also tells us more about a *Burkholderia*–fungus association.

Finally, Partida-Martinez and Hertweck found that although bacterial production of



Figure 1 | Rice under cultivation in Vietnam. World production in 2004 exceeded 600 million tonnes⁷, but rice plants are prey to many diseases, including seedling blight.

P. BRONSTEIN/GETTY



50 YEARS AGO

"An Uncollected Report of the Great Sea-serpent" — Though trivial by itself, it corroborates a number of more detailed accounts of a similar monster in the North Atlantic towards the end of the eighteenth century. Thomas Holcroft...describes an interview on board the *Kennet* (Captain Thompson). After recounting two second-hand stories of the Kraken and comparing these with Pontoppidan's versions, Holcroft continues: "Finding this Leviathan so familiar to their belief, I next inquired if they had heard, or knew any thing of the sea-snake, by some called the sea-worm? To this question I received a still more direct answer. The Mate, Mr. Baird, who certainly was not a liar by habit, whatever mistake or credulity might make him, assured me that, about midway in a voyage to America, in the Atlantic, he had himself seen a fish, comparatively small in the body, of from forty to fifty fathoms in length; and that it had excited great terror in the Captain, who was well acquainted with those latitudes, lest it should sink the ship."

From *Nature* 8 October 1955.

100 YEARS AGO

"Type-Writing by Telegraph" — ...what a type-writing telegraph has to do is the following:—it has to receive a message and translate it into a series of time or magnitude signals; to transmit these signals electrically over a wire, and to re-translate them into a series of space signals.

...There can be no question after the perusal of Mr. Murray's paper that [his system] possesses many advantages over its forerunners which should enable it to survive. It is stated that the automatic part of the apparatus can be run perfectly up to 200 words (1200 letters) a minute, but that no typewriter will stand the strain of being run at this speed, a maximum of 120 words being all that is allowable.

From *Nature* 5 October 1905.

rhizoxin occurs outside the host, it diminishes over time. This may, however, simply be a consequence of an artificial medium that was not entirely favourable to bacterial growth, and does not necessarily mean that the bacteria cannot sustain toxin production in the absence of a host. *Burkholderia* bacteria are known to cause disease in plants², but the fact that they do so in a symbiosis with a fungus is a new finding.

So what's in it for the partners? At this stage, we can only speculate, but there could be several benefits for the bacteria. For example, the fungus may act as a vector for rapid bacterial dispersal to new roots. Further, much plant tissue is difficult to degrade, but fungi are particularly well equipped with the enzymes to do this; presumably, the bacteria reap the benefits of their toxin production when the fungus digests the dead plant cells. For the fungus, toxin release results in a supply of dead organic matter to digest and perhaps also deters other competitors.

More than 125 years ago, biological symbiosis was first defined by Anton de Bary

as simply the cohabitation of two different organisms⁶. Since then it has gradually become clear that the spectrum of relationships varies from beneficial mutualism to outright parasitism. Partida-Martinez and Hertweck's work adds a fresh aspect to our understanding of such relationships — that the apparently beneficial existence of two organisms can occur at the expense of a third one. ■

Ian R. Sanders is in the Department of Ecology and Evolution, Biology Building, University of Lausanne, 1015 Lausanne, Switzerland.

e-mail: ian.sanders@unil.ch

1. Partida-Martinez, L. P. & Hertweck, C. *Nature* **437**, 884–888 (2005).
2. Coeyne, T. & Vandamme, P. *Environ. Microbiol.* **5**, 719–729 (2003).
3. Moran, N. A. & Wernegreen, J. J. *Trends Ecol. Evol.* **15**, 321–326 (2000).
4. Bianciotto, V. et al. *Int. J. Syst. Evol. Microbiol.* **53**, 121–124 (2003).
5. Bianciotto, V. et al. *Appl. Environ. Microbiol.* **70**, 3600–3608 (2004).
6. de Bary, H. A. *Die Erscheinung der Symbiose* (privately printed in Strasbourg, 1879).
7. www.irri.org/science/ricestat/index.asp

CONDENSED-MATTER PHYSICS

Melted by mistakes

Edward J. Kramer

Two-dimensional polymers are potentially useful structures — if we could only understand their properties. Observations of one polymer's intricate, two-stage, melting transition may help us do just that.

Can the polymers from which moulded shoe-soles are made assemble themselves into two-dimensional repeating structures accurately enough to create materials for certain magnetic and semiconductor applications? Such an advance could replace the expensive electron and X-ray-lithography techniques needed for these uses. Writing in *Physical Review Letters*, Angelescu et al.¹ contribute to this vision with experiments that further our understanding of the ordering processes of certain polymers. They also test a theory that the melting of two-dimensional crystals results from heat-induced defects in their structure.

Generally, thin films of block copolymers — polymers comprising blocks of at least two different molecular chains joined covalently — are cast and heated to form a disordered melt monolayer. This monolayer can be cooled to create ordered patterns, on scales of 20–50 nanometres, that can be replicated in underlying layers of inorganic material². In films of the 'A–B diblock' copolymers investigated by Angelescu et al.¹, spherical domains of polymer B are surrounded by continuous domains of polymer A (Fig. 1a). In thick sections, this structure forms a regular array below an order-disorder transition temperature of 121 °C; above this temperature, the packing is liquid-like.

When applied as a 30-nanometre-thick film to a silicon oxide substrate, the polymer diblocks assemble into a hexagonal array of spheres of B above a brush-like layer of A and B.

The melting of analogous two-dimensional arrays consisting of atoms (xenon on graphite, for example), liquid crystalline molecules or colloids has been investigated over the past two decades, driven by the development of the KTHNY theory^{3–5}. This theory — named after the initials of its originators — predicts that the transition from crystal to liquid can be split into two transitions. These transitions, which occur at different temperatures, are associated with the formation of isolated defects known respectively as dislocations and disclinations (Fig. 1b). The intermediate, 'hexatic' state has previously been identified by direct imaging in two-dimensional arrays of block copolymers⁶, magnetic bubbles⁷ and most^{8–10}, but not all¹¹, two-dimensional colloids. In this state, the correlation between the orientations of the 'bonds' of one domain and those of other domains decays only slowly, as a power law, with the distance between the domains. (In the crystal, this orientation correlation does not decay, whereas in the liquid it decays over a few particle diameters.)

In the KTHNY theory, order in the two-

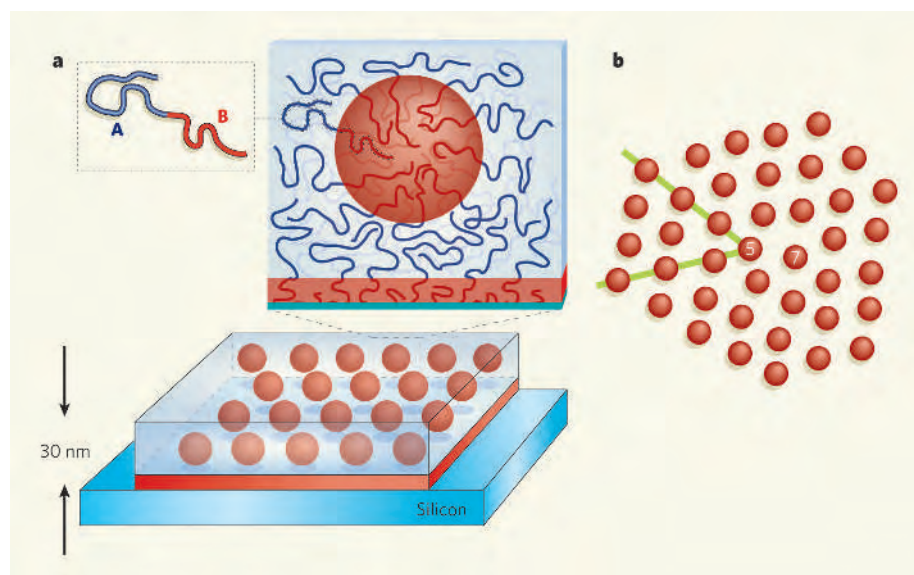


Figure 1 | The A-B diblock copolymer used by Angelescu and colleagues¹. **a**, Each molecule consists of a long block of a polymer A (blue) joined covalently to a short block of polymer B (red). In thick sections, the copolymer self-assembles into a 'body-centred-cubic' structure of spherical domains of B surrounded by a continuous domain of A. In the thin (30 nm) film studied, the B spheres organize themselves into a single layer, hexagonally packed with an average spacing of 25 nm on a 'brush' of A-B chains. **b**, In a perfect two-dimensional hexagonal array, each sphere has six nearest neighbours. A 'dislocation' consists of two extra close-packed rows of spheres (green lines). These terminate at a sphere with only five neighbours, adjacent to a sphere with seven neighbours — a bound pair of 'disclinations'. A sphere with only five nearest neighbours is missing a 60° wedge of spheres (a - 60° disclination); a sphere with seven neighbours has an extra 60° wedge (a + 60° disclination). In the hexatic phase, dislocations are generated thermally, but there are no free disclinations. When the hexatic phase melts, the disclinations separate ('unbind') to form the disordered liquid.

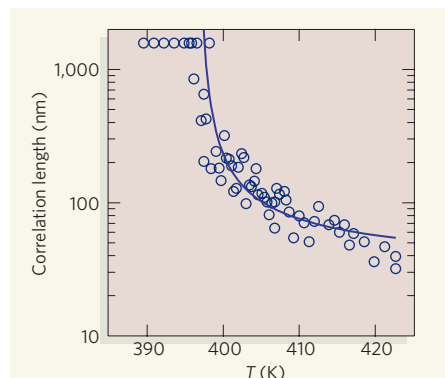


Figure 2 | Good fit? Graph of correlation length ξ against temperature T from data obtained by Angelescu *et al.*¹ (open circles). The constant values at lower T represent lower limits imposed by the finite scan size. The solid line is a fit to the smooth, exponential transition predicted by KTHNY theory: $\xi \sim \exp(B/(T - T_i)^{1/2})$ with the constant $B = 4.28 \text{ K}^{1/2}$ and the transition temperature $T_i = 396.5 \text{ K}$.

dimensional array decreases continuously through both melting transitions — in contrast to the discontinuous transition observed in three-dimensional crystals. But some results from experiments and simulations challenge this prediction. In two-dimensional colloid arrays in which a particle-density gradient has been set up to probe the melting process, both continuous⁸ and discontinuous⁹ transitions have been observed, apparently depending on whether the interaction potential between

the colloidal particles is long- or short-range.

So what of block copolymer spheres? Angelescu *et al.*¹ placed their films in a temperature gradient and used scanning force microscopy to image the degree of order in the array directly, over a range of temperatures that encompassed the hexatic-to-liquid transition. They monitored the changes in the orientation correlation length, ξ — a measure of the distance in the liquid over which the bond orientation around a given sphere is correlated with that of the spheres surrounding it. They found that ξ increases slowly as the temperature, T , decreases towards T_i , the temperature of the hexatic-liquid transition (Fig. 2). As T decreases further, it seems to jump rather abruptly; at the same temperature, the density of disclinations drops to zero.

Such abrupt jumps are not obviously compatible with the smooth increase in order predicted by theory. But KTHNY predicts⁴ that the correlation length increases exponentially with $(T - T_i)^{-1/2}$ as the film is cooled to T_i . This functional form, though continuous, is very different from the power-law divergence that is usually found for continuous phase transitions (for example, the demixing of a two-component fluid). The prediction also fits Angelescu and colleagues' data¹ near the transition to within the experimental scatter (Fig. 2). So it seems that this aspect of the KTHNY theory escapes its nemesis for now. Given the 'hockey-stick' form of the increase in KTHNY correlation length near the transition, it will be

difficult to devise an experiment that would conclusively support either the abrupt-jump or exponential interpretation.

This experimental ambiguity should not obscure the fact that two-dimensional block copolymer layers show clear evidence of a hexatic phase between the solid and liquid states — an essential prediction of the KTHNY theory. Crucially, the existence of the hexatic phase is likely to lead to new methods for preparing block copolymer monolayers with better order: the straight edges of channels confining a monolayer induce an isotropic liquid to freeze to the hexatic state with a well-defined orientation near the edge¹².

Other types of block copolymer monolayers also melt differently from three-dimensional crystals. Cylinders lying parallel to the substrate, for example, are predicted¹³ to experience the thermal generation of dislocations, whose disclinations also separate from each other on melting, a prediction recently verified experimentally¹⁴. No experimental evidence is yet available, however, on whether the melting transition to the isotropic liquid is continuous, as predicted by theory.

The mechanisms of defect removal as the copolymers cool may dictate whether two-dimensional block copolymer films can be made with sufficient control over their patterning to be useful for lithography. The results of Angelescu *et al.*¹ highlight the importance of eliminating thermally generated defects during the cooling process. Their experiments¹⁵, as well as simulations¹⁶, have already provided valuable insights into the mechanisms controlling this kinetics. But even when expensive lithography is required, two-dimensional block copolymers could play a critical role in smoothing the roughness of pattern edges¹⁷.

Edward J. Kramer is in the Departments of Materials and Chemical Engineering, University of California II-1361C, Santa Barbara, California 93106, USA. e-mail: edkramer@mrl.ucsb.edu

1. Angelescu, D. A., Harrison, C. K., Trawick, M. L., Register, R. A. & Chaikin, P. M. *Phys. Rev. Lett.* **95**, 025702 (2005).
2. Segalman, R. A. *Mater. Sci. Eng. R* **48**, 191-226 (2005).
3. Kosterlitz, J. M. & Thouless, D. J. *J. Phys. C* **6**, 1181-1203 (1973).
4. Nelson, D. R. & Halperin, B. I. *Phys. Rev. B* **19**, 2457-2484 (1979).
5. Young, A. P. *Phys. Rev. B* **19**, 1855-1866 (1979).
6. Segalman, R. A., Hexemer, A., Hayward, R. C. & Kramer, E. J. *Macromolecules* **36**, 3272-3288 (2003).
7. Seshadri, R. & Westervelt, R. M. *Phys. Rev. B* **46**, 5150-5161 (1992).
8. Murray, C. A., Sprenger, W. O. & Wenk, R. A. *Phys. Rev. B* **42**, 688-703 (1990).
9. Marcus, A. H. & Rice, S. A. *Phys. Rev. E* **55**, 637-656 (1997).
10. Zahn, K., Lenke, R. & Maret, G. *Phys. Rev. Lett.* **82**, 2721-2724 (1999).
11. Karnchanaphanurach, P., Lin, B. H. & Rice, S. A. *Phys. Rev. E* **61**, 4036-4044 (2000).
12. Segalman, R. A., Hexemer, A. & Kramer, E. J. *Phys. Rev. Lett.* **91**, 196101 (2003).
13. Toner, J. & Nelson, D. R. *Phys. Rev. B* **23**, 316-334 (1981).
14. Hammond, M. R., Cochran, E., Fredrickson, G. H. & Kramer, E. J. *Macromolecules* **38**, 6575-6585 (2005).
15. Harrison, C. *et al. Europhys. Lett.* **67**, 800-806 (2004).
16. Vega, D. A. *et al. Phys. Rev. E* **71**, 061803 (2005).
17. Stoykovich, M. P. *et al. Science* **308**, 1442-1446 (2005).

ECOLOGY

Stars beneath the waves

Astronomy and population ecology seemingly have little in common. But they meet in a paper by Zaven Arzoumanian and colleagues, who have developed a way of recognizing individual spot patterns of whale sharks by adapting an algorithm used for comparing star patterns in images of the night sky (*J. Appl. Ecol.* doi:10.1111/j.1365-2664.2005.01117.x).

Many animals have beautiful spot patterns that differ subtly between individuals of the same species. Photographs of these markings allow ecologists to identify individuals in population studies, but sifting through them is laborious and often unreliable when done by eye. The aim of the authors — an astronomer, a computer scientist and a marine biologist — was to automate the process.

The whale shark (*Rhincodon typus*, pictured) is the world's largest fish, reaching lengths of more than 14 metres, and bears patterns of white spots on its dark dorsal surface. The leap forwards made by

Arzoumanian *et al.* was to view these patterns as celestial constellations. They adapted an algorithm for identifying characteristic stellar patterns from the geometric properties of triangles made by the lines formed between all possible combinations of three-spot coordinates within a defined region of space. In the same way, by concentrating on a highly variable region of spots above the pectoral fin of whale sharks, the spatial relationships between spots formed the basis for identifying a shark's unique spot pattern.

The modified algorithm was used to calculate similarity scores by comparing the geometry of each spot triangle in a shark photograph with all spot-triangle combinations from a second image. This procedure was repeated between all image-pair combinations to locate high-scoring matches. The researchers then tested the technique by scanning a web-based database of around 1,500 whale shark photographs, and achieved a 92% success rate in



J.D. WATT/IMAGE QUEST 3-D

matching previously identified pairs of images of individual sharks, as well as resolving new matches not detected by eye. The approach should improve our understanding of whale shark population dynamics, about which little is known. More information is needed urgently, as increasing exploitation of whale sharks means their populations are now vulnerable.

Many animal species have star-quality skin patterns, and the study shows how individual animals can be 'tagged' at low cost for long

periods, and without physical interference. Linking the pattern-matching algorithm to a web-based photo-identification database, so that it performs like an Internet search engine, will allow researchers around the world to compare new images with vast numbers of virtually tagged animals.

David W. Sims

David W. Sims is at the Marine Biological Association of the United Kingdom, The Laboratory, Citadel Hill, Plymouth PL1 2PB, UK. e-mail: dws@mba.ac.uk

PALAEOBIOLOGY

Sea change in sediments

David J. Des Marais

Earth's oxygen levels increased slowly over a long and ill-defined transitional period around two billion years ago. A microbial 'footprint' from this era provides biological evidence to complement existing geological data.

The microorganisms that were the sole forms of life on the early Earth survived billions of years of profound environmental change. Their modern counterparts exhibit a broad range of metabolic capabilities, exploiting diverse energy sources and so thriving in many different environments. But how did this remarkable adaptability develop? And have microorganisms preserved a record of past environmental changes? In the absence of complex body fossils laid down by microorganisms, Brocks and colleagues (page 866 of this issue)¹ explore the early biosphere by analysing the chemical 'footprint' left by microorganisms on an ocean floor some 1.6 billion years ago, and compare it with how such a footprint would look today. In doing so, they confirm geochemical evidence of changes in the composition of Earth's oceans and

atmosphere that occurred over a long period around two billion years ago.

In the primordial Archaean era, Earth's oceans and atmosphere contained little or no free oxygen. In this anoxic era, reduced compounds (the chemical opposites of oxidized compounds, containing species with an increased number of electrons), such as sulphides and ferrous-iron compounds, persisted for longer than they do in today's well-oxygenated marine and land environments^{2,3}. Anaerobic photosynthesizing microorganisms used these reduced inorganic compounds to harvest energy for the synthesis of their cellular constituents. This situation began to change with the emergence of oxygen-producing photosynthetic cyanobacteria around 2.7 billion years ago, perhaps earlier (Fig. 1).

Photosynthesis is used by at least five phototrophic bacterial groups. In four of these groups — green sulphur bacteria, green non-sulphur bacteria, heliobacteria and proteobacteria — the process is anoxygenic (no oxygen is produced). These microbiota also require a reduced chemical substrate to provide electrons for organic biosynthesis⁴. Cyanobacteria, which form the fifth group, are unique in harnessing light directly to split water molecules, releasing molecular oxygen as a by-product. Oxygenic photosynthesis liberated emergent microbial life from its total dependence on hydrothermal fluids and other sources of reduced inorganic compounds; cyanobacteria could then spread across the planet.

A significant initial effect of oxygenic photosynthesis on the oceans was the oxidation of reduced compounds containing ferrous iron and sulphides — at least in surface waters. Gradually, oxygen levels increased until the oceans resembled the well-oxygenated seas we know today. Evidence that this transition was slow has previously been garnered from geological and geochemical features, such as the large bodies of iron-rich sedimentary rocks known as banded iron formations. Ferrous iron supplied through weathering and hydrothermal activity was deposited in these

layers for hundreds of millions of years until the process came to an abrupt halt² around 1.7 billion years ago, when the presence of sufficient photosynthetic oxygen oxidized ferrous iron in sea water. Sometime later, sulphides became rapidly oxidized during weathering, and sulphates, their oxidized products, accumulated in sea water⁵.

Brocks and colleagues¹ pursue the crucial question of exactly when marine oxygen levels approached modern values — and when, for example, communities of the aerobic planktonic, eukaryotic (nucleus-containing) cellular organisms characteristic of today's oceans became established. They investigated fossilized hydrocarbons in siltstones and shales in a former marine sedimentary basin in northern Australia dating from the mid-Proterozoic era, 1.64 billion years ago. The synthesis of certain hydrocarbons is specific to certain groups of microorganisms⁶, so the fossilized hydrocarbons present in the rocks act as 'biomarkers' for the bacteria active at the time of their deposition.

By comparing these biomarkers with those from analogous modern environments¹, a portrait can be painted of a global environment 1.64 billion years ago that was at a degree of oxidation intermediate between those of the Archaean and modern worlds. The presence of β -carotane in the record, for example, probably signals that cyanobacteria supplying reduced carbon were present. Aerobic methane-oxidizing bacteria, which flourish in environments where methane and oxygen are available but oxygen levels are low, were seemingly also abundant. This methane was produced by a group of microorganisms known as archaea that must have competed successfully with sulphate-reducing bacteria for reduced chemical substrates; this indicates that sulphate levels were substantially lower than in modern sea water. On the other hand, the sulphate supply to sulphate-reducing bacteria was at least sufficient for these microorganisms to produce some sulphides.

The carotenoids found by Brocks and colleagues¹ in their rock samples speak for the presence of non-oxygenic photosynthetic bacteria of the type that typically requires both reduced sulphur compounds and sunlight for growth — so indicating the presence of shallow, sunlit sulphidic waters lacking oxygen. The abundances of pyrite (iron sulphide, FeS_2) and stable sulphur isotopes in mid-Proterozoic sediments^{5,7} support the inference that sulphate levels were substantially lower than those today, and that sulphide was pervasive in the water column. Finally, the conspicuous absence of eukaryotic biomarkers in the samples seems to indicate that the pervasive planktonic populations so characteristic of later marine environments were not to be found.

So why did atmospheric and marine oxygen fail to attain near-modern levels during the

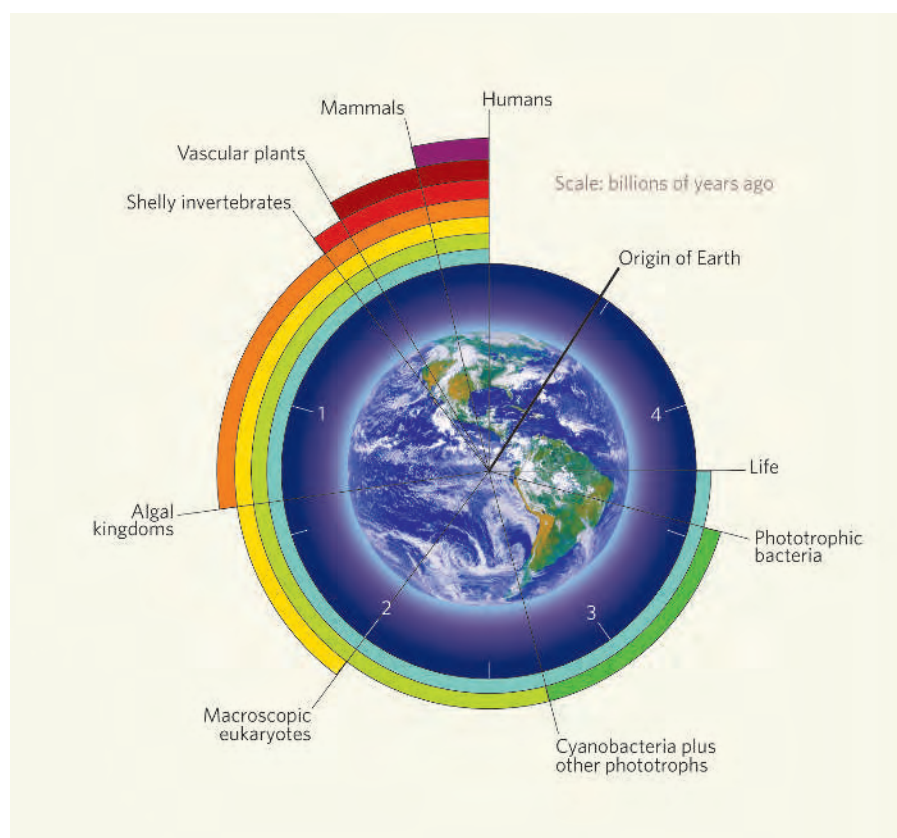


Figure 1 | Life on Earth through the ages. The ages of earliest known evidence of various biota (shown here) are probably determined by the quality of preservation of such ancient evidence. Bacteria capable of photosynthesis developed sometime before 2.7 billion years ago. The oxygen-producing cyanobacteria started the prolonged process of oxygenation of Earth's atmosphere and oceans that ultimately allowed more complex aerobic organisms to develop. Brocks and colleagues' analyses¹ of biomarkers for microorganisms in marine sediments 1.6 billion years old provide a snapshot of an ocean in an intermediate stage of oxygenation: despite the presence of cyanobacteria, the waters were largely anoxic and sulphidic, with no evidence for the aerobic, planktonic eukaryotes characteristic of the modern ocean.

mid-Proterozoic? A possible answer is that biologically important metals such as iron and molybdenum were stripped from sea water to form highly insoluble sulphides⁸. Sulphidic waters are toxic to planktonic eukaryotes and therefore possibly suppressed their populations. This mid-Proterozoic period of deep-sea anoxia and restricted productivity probably ended when increased rates of oxidative weathering of sulphides accelerated sulphate production and caused the sulphidic waters to retreat⁵.

Brocks and colleagues' study¹ greatly extends the known antiquity of key biomarkers and is a significant contribution to charting the evolution of the mid-Proterozoic biosphere. But it is also just a first step. Now these microbial biomarkers should be analysed for their carbon-isotopic compositions to help confirm the microorganisms' identities and clarify their relationships within their communities⁹. Biomarkers should be characterized in mid-Proterozoic sedimentary rocks along an onshore-offshore axis to reconstruct the coastal palaeogeography more completely. Further analyses of biomarkers in rocks of similar age from other localities should help

to provide a more global perspective. And sediments of various ages should be examined to determine, among other things, how the distribution of eukaryotic biomarkers might be related to changes in the biogeochemical sulphur cycle. Such studies will indeed add pieces to the puzzle of the microbial populations and processes that eventually led to the modern oxygenated biosphere. ■

David J. Des Marais is in the Exobiology Branch and the NASA Astrobiology Institute, MS 239-4, NASA Ames Research Center, Moffett Field, California 94035-1000, USA.
e-mail: david.j.desmarais@nasa.gov

1. Brocks, J. J. *et al.* *Nature* **437**, 866–870 (2005).
2. Holland, H. D. *The Chemical Evolution of the Atmosphere and Oceans* (Princeton Univ. Press, 1984).
3. Farquhar, J., Bao, H. & Thieme, M. *Science* **289**, 756–758 (2000).
4. Blankenship, R. E. *Photosynth. Res.* **33**, 91–111 (1992).
5. Canfield, D. E. *Nature* **396**, 450–453 (1998).
6. Brocks, J. J. & Summons, R. E. in *Treatise on Geochemistry* Vol. 8 (ed. Schlesinger, W. H.) 63–115 (Elsevier, Amsterdam, 2004).
7. Kah, L. C., Lyons, T. W. & Frank, T. D. *Nature* **431**, 834–838 (2004).
8. Anbar, A. D. & Knoll, A. H. *Science* **297**, 1137–1142 (2002).
9. Hayes, J. M. *Rev. Mineral.* **43**, 225–277 (2001).

OBITUARY

Hermann Bondi (1919–2005)

Mathematician, cosmologist and public servant.

Hermann Bondi, who died on 10 September aged 85, is perhaps best known as one of the original advocates of the steady-state theory of the Universe, advanced as an alternative to the now dominant Big Bang theory. But as an astrophysicist, mathematician and cosmologist, his interests were wide-ranging and his intellectual brilliance undoubted. As a scientific adviser to successive British governments, he also contributed immensely to the public life of his adopted country.

Bondi was born in Austria to Jewish parents. As a schoolboy in Vienna he showed precocious mathematical ability, and decided that he wanted to work with the renowned mathematician and astronomer Arthur Eddington at the University of Cambridge. Bondi arrived at the university's Trinity College in 1937. In March 1938, German troops entered Austria and *Anschluss* was proclaimed. Bondi's parents and sister followed his cabled advice to "drop everything and leave", settling eventually in New York; he himself, an anglophile, remained based in England for the rest of his life.

In the panic following the collapse of France in 1940, Bondi was interned on the Isle of Man and then in Canada. With typical resourcefulness, he helped set up a camp university at which he gave lectures on mathematics. In Canada, he met a fellow Viennese internee, Thomas Gold, who became his lifelong friend and collaborator. On their return to Britain in 1941, Bondi and Gold joined the Admiralty's radar research group, led by Fred Hoyle. It was this trio's off-duty discussions on astronomy that were to bear fruit when they returned to Cambridge at the end of the war.

This was a time of momentous advances in astrophysics. In the late 1930s, Hans Bethe and Carl-Friedrich von Weizsäcker had shown that, at the high temperatures inside stars, energy liberated from the nuclear fusion of hydrogen to form helium could supply stellar luminosity. This led Hoyle and Ray Lyttleton to a reappraisal of the physics and mathematics of stellar structure developed by Bondi's hero Eddington. Together with his wife Christine — one of Hoyle's doctoral students — Bondi transformed these equations, so simplifying the construction of stellar models, especially those of chemically inhomogeneous stars. Bondi's student, Roger Tayler, exploited their technique further in a thesis that foreshadowed the now established picture of stellar evolution through nuclear processing from the main sequence to the giant branch.

Bondi also applied his mathematical skill

to various problems in cosmic gas dynamics. In particular, he wrote an elegant paper on the spherically symmetric, pressure-limited gravitational accretion of interstellar gas by a star at rest in a gas cloud. An earlier paper, written with Hoyle and Lyttleton, had argued that gravitational energy released by accreted gas heats the corona of the Sun to more than a million degrees. Although coronal heating is now known to come from the dissipation of energy propagating outwards from the Sun, the resulting coronal expansion — the 'solar wind' — is most simply described by changing the boundary conditions in Bondi's spherical accretion model to yield an analogous flow with opposite sign.

But Bondi is undoubtedly best known for his partnership with Hoyle and Gold in the advocacy and detailed study of steady-state cosmology. In this model, the expansion of the Universe — inferred by Edwin Hubble from the redshifted spectral lines of distant galaxies — is exactly compensated by the continuous creation of matter. According to this idea, the mean density of matter in the cosmos remains constant; the Universe has no beginning, no end and no age. Bondi's motivation came in part from observational difficulties then besetting the rival Big Bang cosmology (to use Hoyle's intentionally pejorative term), but also from his conviction that an acceptable cosmology must satisfy Mach's principle, which relates local inertial effects to the global properties of the cosmos.

His philosophy of science was strongly influenced by Karl Popper. He favoured steady-state cosmology because it was 'vulnerable' — it made precise predictions that were falsifiable, in Popper's sense of the word. This puritanical stance was tempered by his appreciation that incontrovertible observational facts were often hard to find. Thus he pointed out that the mass of Capella, used by Eddington to calibrate his stellar mass–luminosity relation, had since been revised. Equally, one 'fact' that cast doubt on evolutionary cosmology — the ages of the oldest stars apparently exceeded the estimated age of the Universe — disappeared in 1952 with Walter Baade's doubling of the cosmic distance scale.

Bondi remained true to Popper, and gave up the steady-state hypothesis in the face of increasing conflicting observational evidence, such as the discovery in 1965 of the cosmic microwave background radiation, which the steady-state hypothesis could not easily explain. He showed little interest in the subsequent revised versions proposed by



Hoyle and his collaborators. Bondi's legacy in this area is his superbly written monograph *Cosmology*, and a series of papers on general relativity, especially on gravitational radiation. The list of concepts that bear his name shows his influence: the Bondi news function, the Bondi mass, Bondi waves, the Bondi–Metzner–Sachs group and the Tolman–Bondi universe.

From 1951, Bondi's academic career continued at King's College London, and culminated in his election in 1983 as master of Churchill College, Cambridge. But he also began a parallel life as a high-level public servant, becoming director-general of the forerunner of the European Space Agency, the Paris-based European Space Research Organization, in 1967. In a subsequent wide-ranging career in Britain that was testament to both his intellectual and his organizational abilities, he was successively chief scientific adviser to the Ministry of Defence, chief scientist to the Department of Energy, where he laid the groundwork for Britain's first long-term energy policy, and chairman of the Natural Environment Research Council. One of his greatest achievements was his report written after the 1953 floods that devastated parts of eastern England, which ultimately led to the construction of the Thames Barrier to protect London from storm surges.

Bondi believed strongly in the social responsibility of scientists, and so was active in the Pugwash movement and the scientific education of the public. From 1982 to 1999 he was president of the British Humanist Association. Although he could be ruthless when confronted with flawed arguments, whatever their source, he respected the intellectually honest opponent working from different axioms. He will be much missed, as both a public servant and a private man. ■

Leon Mestel

Leon Mestel is emeritus professor of astronomy at the University of Sussex, Falmer, Brighton, East Sussex BN1 9QH, UK.
e-mail: lmestel@sussex.ac.uk

P. YEMPICS

BRIEF COMMUNICATIONS

Asian honeybees parasitize the future dead

When a queen dies, unrelated workers seize the chance to move into her nest and lay their own eggs.

The queen of a honeybee colony has a reproductive monopoly because her workers' ovaries are normally inactive and any eggs that they do lay are eaten by their fellow workers^{1–3}. But if a colony becomes queenless, the workers start to lay eggs, stop policing² and rear a last batch of males before the colony finally dies out⁴. Here we show that workers of the Asian dwarf red honeybee *Apis florea* from other colonies exploit this interval as an opportunity to move in and lay their own eggs while no policing is in force. Such parasitism of queenless colonies does not occur in the western honeybee *A. mellifera* and may be facilitated by the accessibility of *A. florea* nests, which are built out in the open.

Apis florea are small honeybees that nest by building a single comb attached to a twig (Fig. 1). We collected four wild *A. florea* nests (one in 2003 and three in 2004) and transported them to a different location that hosts many wild *A. florea* colonies, tying them on to low tree branches at least 5 m away from any other nest. After taking a sample of workers, we removed the queens from the translocated colonies and also any queen cells that subsequently developed. We sampled adult workers after one week and again after four weeks (pooled data are shown in Table 1). Samples of worker-produced eggs, larvae and pupae were collected as they appeared in the combs (but

not for colonies 3 and 4, as these absconded before larvae were reared). We dissected the adult workers to determine their ovary activation³ and used analysis of DNA microsatellite loci to determine their parentage and that of worker-produced males⁵. (For details, see supplementary information.)

Before the queen was removed, the number of unrelated (non-natal) workers in the colony was low (averaging 2.0%; Table 1) and none of these workers had activated ovaries. After queen removal, however, the proportion of non-natal workers rose significantly ($P = 0.008$) to 4.5%. Significantly more ($P < 0.001$) non-natal workers (42.6%) had activated ovaries than did natal workers (17.7%), indicating that parasitic workers may actively seek out queenless colonies in order to lay eggs. Moreover, non-natal workers had significantly higher reproductive success ($P < 0.001$) than natal workers: 3.2% of workers in colonies 1 and 2 were non-natal, but these laid 35.6% of the eggs and 22.5% of the pupae.

These results show that an important reproductive tactic of *A. florea* workers is actively to seek out and parasitize queenless nests with their eggs. This behaviour is not evident in the western honeybee, *A. mellifera*, in which the offspring of non-natal workers are rare or absent in queenless nests⁵. Our findings could also explain why queenless dwarf-bee colonies often abscond from their nests: these workers opt to join new, unrelated nests and parasitize them with their eggs. Even workers from



Figure 1 | Out in the open. The nest of the Asian dwarf red honeybee, *Apis florea*, is built as a single comb suspended from a twig. This makes it accessible to workers invading from other colonies when the queen dies.

Table 1 | Reproductive parasitism of queenless colonies

	Colony			
	1	2	3	4
Before queen removal				
Number of non-natal workers	1	5	0	2
Number of natal workers	99	95	96	94
After queen removal				
Number of non-natal workers	0	13	8	14
Number of natal workers	200	187	184	178
Non-natals with activated ovaries (%)	–	15.4	62.5	50.0
Natals with activated ovaries (%)	28.0	27.8	4.9	10.1
Offspring derived from non-natal workers (%) (number of offspring sampled)				
Eggs	44.4 (115)	26.7 (120)	–	–
Larvae	38.5 (143)	25.6 (125)	–	–
Pupae	30.0 (100)	14.9 (94)	–	–

A. florea colonies that have a queen may choose to parasitize queenless nests, perhaps because they favour individual reproduction in a queenless nest over contributing to the reproductive output of their own colony.

Piyamas Nanork*†‡, Jürgen Paar†, Nadine C. Chapman†, Siriwat Wongsiri*, Benjamin P. Oldroyd†

*Department of Biology, Chulalongkorn University, Bangkok 10330, Thailand

†School of Biological Sciences A12, University of Sydney, New South Wales 2006, Australia

e-mail: boldroyd@bio.usyd.edu.au

‡Department of Biology, Mahasarakham University, Mahasarakham 44150, Thailand

1. Ratnieks, F. L. W. & Visscher, P. K. *Nature* **342**, 796–797 (1989).
2. Miller, D. G. & Ratnieks, F. L. W. *Insectes Sociaux* **48**, 178–184 (2001).
3. Halling, L. A. et al. *Behav. Ecol. Sociobiol.* **49**, 509–513 (2001).
4. Page, R. E. & Robinson, G. E. *Behav. Ecol. Sociobiol.* **35**, 99–107 (1994).
5. Martin, C. J., Oldroyd, B. P. & Beekman, M. *Behav. Ecol. Sociobiol.* **56**, 42–49 (2004).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.
doi:10.1038/437829a

SEISMOLOGY

Dynamic triggering of earthquakes

After an earthquake, numerous smaller shocks are triggered over distances comparable to the dimensions of the mainshock fault rupture, although they are rare at larger distances. Here we analyse the scaling of dynamic deformations (the stresses and strains associated with seismic waves) with distance from, and magnitude of, their triggering earthquake, and show that they can cause further earthquakes at any distance if their amplitude exceeds several microstrain, regardless of their frequency content. These triggering requirements are remarkably similar to those measured in the laboratory for inducing dynamic elastic nonlinear behaviour, which suggests that the underlying physics is similar^{1,2}.

We assume that some aftershocks are dynamically triggered³ and therefore that, within aftershock zones, mainshock-generated dynamic deformations must be sufficiently large to trigger earthquakes. The square root of the rupture area, $\sqrt{\Sigma}$, provides a useful approximate measure of both rupture dimension and the aftershock zone⁴. We measured the horizontal component of the seismic waves' peak ground-motion velocity (PGV, which approximates the peak shear strain when divided by the shear wave or phase velocity⁵), recorded at distances of 0.1 km to 5,300 km from earthquakes

with magnitudes, M , of 4.4 to 7.9 (Fig. 1a).

To define all the aftershock zones by the same scaled distance, for each earthquake we divide the distances corresponding to each PGV measurement by an estimate of $\sqrt{\Sigma}$ (from the database of Finite-Source Rupture Models at www.seismo.ethz.ch/srcmod). The scaling by $\sqrt{\Sigma}$ eliminates any resolvable dependence on earthquake size: this can be explained by a model of seismic radiation from a fault with area Σ . Thus, the PGVs at a scaled distance around unity define an approximate minimum triggering threshold of several to ten microstrain (Fig. 1b).

Remote, spatially widespread triggering has been most clearly observed following the earthquakes at Landers⁶ ($M=7.3$) and Denali⁷ ($M=7.9$) in California and Alaska, respectively, and (although less pronounced) after the Hector Mine earthquake⁶ ($M=7.1$) in California. PGVs for these seem consistent with the inferred threshold (Fig. 1b) at all but one site (the geothermal area of Long Valley, California). Landers data exist at very remote distances only for non-triggered sites, but these provide a lower bound on triggering deformations as they are off the azimuth of expected focusing. The absence of triggering where strain amplitudes exceed the proposed

threshold implies that large deformations may be a necessary but not sufficient condition.

These observations indicate that remote triggering may require exceptionally large dynamic deformations, perhaps as a result of strong directivity^{6,7}, thereby explaining why this occurs only rarely. That a simple amplitude threshold seems to account for both the occurrence and absence of triggering, and the fact that the PGVs come from signals with very different frequency contents (dominant frequencies are roughly proportional to Σ), also implies that the mechanisms of dynamic triggering do not depend strongly on frequency.

We have proposed a model in which dynamic deformations promote earthquake failure by mechanisms involving dynamic nonlinear elasticity and slow dynamics¹. We base this on the similarities between our seismological and laboratory observations¹ (Fig. 1b, and see supplementary information), field observations⁸, and on the modelling² of dynamic nonlinear elasticity. Application of dynamic strains of the order of several microstrain seems to be required both for dynamic triggering of earthquakes and for the significant nonlinearity that arises from modulus reduction (softening) in laboratory and field experiments and in models^{1,2}. Another similarity is a lack of dependence on loading frequency over bandwidths spanning several orders of magnitude². If a fault is in a critical state near to failure, we suggest that softening leads to failure^{1,9}.

Although we have not considered the extended durations of triggered earthquake sequences, our model explains them through the recovery that follows dynamic softening by waves (that is, the slow dynamics) from both the mainshock and from creep following subsequent, locally triggered earthquakes¹. We should be able to validate this model as new earthquake and lab data become available.

Joan Gomberg*, Paul Johnson†

*US Geological Survey, Suite 2, Memphis, Tennessee 38152, USA
e-mail: gomberg@usgs.gov

†Geophysics Group EES-11, Los Alamos National Laboratory of the University of California, MS D443, Los Alamos, New Mexico 87545, USA

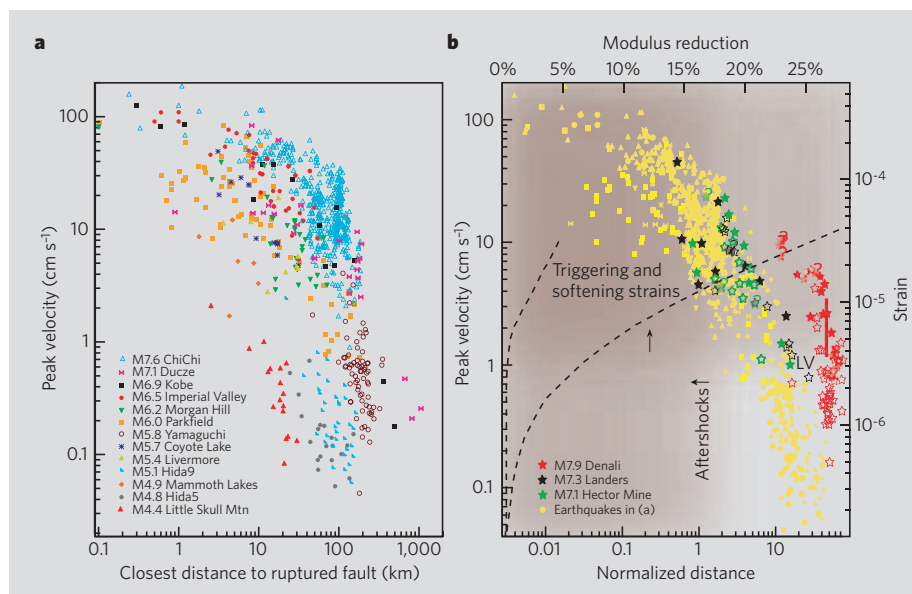


Figure 1 | Dynamic deformation scaling. **a**, Plot of measured peak ground velocities (PGVs) against distance (data from Pacific Earthquake Engineering Research Center; Institutions for Research in Seismology; Japan's National Research Institute for Earth Science and Disaster Prevention K-Net; and Seismology Lab, University of Nevada, Reno). **b**, Plot of measurements in **a** against distance, normalized by rupture dimensions, showing almost identical scaling. Deformations at normalized distances of less than about 1 (left, light brown) must be sufficient to trigger aftershocks; triggering strains lie above the smallest PGV in this normalized-distance range (top, darker brown). This triggering threshold range is consistent with PGVs for three earthquakes that triggered seismicity remotely, measured at sites that did (filled stars) and did not (open stars) experience triggered seismicity (light shading, ambiguous observations). Red bar, additional Denali triggering PGVs¹⁰; dashed curves bound laboratory measurements of modulus reduction against dynamic loading strain amplitude for various rock types, pressures and saturations; LV, Long Valley.

- Johnson, P. A. & Jia, X. *Nature* **437**, 871–874 (2005).
- Guyot, R. A. & Johnson, P. A. *Physics Today* **52**, 30–35 (1999).
- Gomberg, J., Bodin, P. & Reasenberg, P. A. *Bull. Seismol. Soc. Am.* **93**, 118–138 (2003).
- Kanamori, H. & Brodsky, E. E. in *Reports on Progress in Physics* 1429–1496 (Institute of Physics, UK, 2004).
- Gomberg, J. & Agnew, D. J. *Geophys. Res.* **86**, 212–220 (1996).
- Gomberg, J., Reasenberg, P. A., Bodin, P. & Harris, R. A. *Nature* **411**, 462–466 (2001).
- Gomberg, J., Bodin, P., Larson, K. & Dragert, H. *Nature* **427**, 621–624 (2004).
- Field, E. H., Johnson, P. A., Beresnev, I. & Zeng, Y. *Nature* **390**, 599–602 (1997).
- Rudnicki, J. W. *J. Geophys. Res.* **82**, 844–854 (1977).
- Pankow, K. L., Arabasz, W. J., Pechmann, J. C. & Nava, S. J. *Bull. Seismol. Soc. Am.* **94** (suppl.), 332–347 (2005).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.
doi:10.1038/nature04167

Structure of the CED-4–CED-9 complex provides insights into programmed cell death in *Caenorhabditis elegans*

Nieng Yan¹, Jijie Chai¹, Eui Seung Lee^{2,3}, Lichuan Gu¹, Qun Liu⁴, Jiaqing He⁵, Jia-Wei Wu¹, David Kokel², Huilin Li⁵, Quan Hao⁴, Ding Xue² & Yigong Shi¹

Interplay among four genes—*egl-1*, *ced-9*, *ced-4* and *ced-3*—controls the onset of programmed cell death in the nematode *Caenorhabditis elegans*. Activation of the cell-killing protease CED-3 requires CED-4. However, CED-4 is constitutively inhibited by CED-9 until its release by EGL-1. Here we report the crystal structure of the CED-4–CED-9 complex at 2.6 Å resolution, and a complete reconstitution of the CED-3 activation pathway using homogeneous proteins of CED-4, CED-9 and EGL-1. One molecule of CED-9 binds to an asymmetric dimer of CED-4, but specifically recognizes only one of the two CED-4 molecules. This specific interaction prevents CED-4 from activating CED-3. EGL-1 binding induces pronounced conformational changes in CED-9 that result in the dissociation of the CED-4 dimer from CED-9. The released CED-4 dimer further dimerizes to form a tetramer, which facilitates the autoactivation of CED-3. Together, our studies provide important insights into the regulation of cell death activation in *C. elegans*.

Genetic analyses in *C. elegans* led to the identification of four genes—*egl-1* (*egl*, for egg-laying defective), *ced-9* (*ced*, for cell-death abnormal), *ced-4* and *ced-3*—that collectively control the death of 131 somatic cells during hermaphrodite development^{1,2} (Fig. 1a). The protein products of these four genes define a linear pathway. CED-3 is a caspase^{3,4}: a cysteine-containing protease that cleaves its substrates after aspartate residues⁵. CED-3 is synthesized as an inactive zymogen. When cells are programmed to die, the CED-3 zymogen is thought to be activated by the adaptor molecule CED-4 (refs 6–11). In healthy cells, the pro-apoptotic protein CED-4 is sequestered by the mitochondria-bound protein CED-9 (refs 7, 10, 12–15), and thus is unable to activate CED-3. At the onset of cell death, the inhibitory CED-4–CED-9 interaction is disrupted by the pro-apoptotic protein EGL-1 (refs 16–19), which is transcriptionally activated in cells destined to die. The released CED-4 is thought to undergo homooligomerization, which then facilitates the activation of the CED-3 caspase¹¹.

The sequestration of CED-4 by CED-9 is essential for the regulation of programmed cell death in nematodes (Fig. 1a). However, it is not yet understood how CED-9 specifically recognizes the pro-apoptotic protein CED-4. It is also unclear how this recognition results in the inability of CED-4 to activate CED-3. In addition, despite some biochemical evidence²⁰, the mechanism by which EGL-1 releases CED-4 from the inhibitory CED-4–CED-9 complex remains unresolved. Last, but not least, the genetically identified pathway of CED-3 activation (Fig. 1a) has not been fully reconstituted *in vitro* using homogeneous recombinant proteins. In this study, we provide answers to these important questions using an integrated approach of structural biology, biochemistry, biophysics and genetics.

Structure of the CED-4–CED-9 complex

Assuming a 1:1 stoichiometry, the CED-4–CED-9 complex is predicted to exhibit a molecular weight of ~90 kDa. However, the apparent molecular weight of the complex always seemed to be ~150 kDa by gel filtration (see below). This complex, containing the full-length CED-4 and a transmembrane-segment-deleted CED-9 (residues 48–251), was crystallized. The structure was determined by multiple-wavelength anomalous dispersion (MAD) and refined at 2.6 Å resolution (Fig. 1b, Supplementary Table 1).

The structure shows an unexpected 2:1 stoichiometry between CED-4 and CED-9, accounting for the observed mass of ~150 kDa. The two molecules of CED-4, referred to as CED-4a and CED-4b, form an asymmetric dimer through an extensive interface involving a buried surface area of 4,167 Å² (Fig. 1b, c). Each CED-4 protein contains a bound ATP molecule and an associated magnesium ion. Only one CED-4 molecule, CED-4a, directly interacts with CED-9. CED-4a exhibits an elongated shape, with CED-9 and CED-4b binding to opposite sides (Fig. 1c). The carboxy-terminal helix of CED-9 protrudes from the CED-4–CED-9 complex; presumably, this arrangement allows convenient access of the CED-9 C-terminal transmembrane region to the mitochondrial outer membrane.

CED-4a comprises twenty-five α-helices and eight β-strands, which are organized into four sequential domains: an amino-terminal caspase recruitment domain (CARD; residues 1–105), a three-layered α/β-fold characteristic of the P-loop NTPases (residues 106–290), a helical domain (residues 291–370) and an extended winged-helix domain (residues 371–549) at the C terminus (Fig. 1 and Supplementary Fig. 1a). These four domains stack up closely against one another. CED-4b exhibits a similar structure, except that its CARD domain is flexible and disordered in the crystals.

¹Department of Molecular Biology, Princeton University, Lewis Thomas Laboratory, Washington Road, Princeton, New Jersey 08544, USA. ²Department of Molecular, Cellular and Developmental Biology, University of Colorado, Boulder, Colorado 80309, USA. ³Department of Life Science, Gwangju Institute of Science and Technology, Gwangju 500-712, Republic of Korea. ⁴Molecular Biology and Genetics, Cornell University, Ithaca, New York 14853, USA. ⁵Biology Department, Brookhaven National Laboratory, Upton, New York 11973, USA.

The structure of CED-4 within the inhibitory CED-4–CED-9 complex exhibits a different domain organization compared to that of the autoinhibited apoptotic protease-activating factor 1 (Apaf-1; ref. 21), the mammalian homologue of CED-4 (ref. 22) (Supplementary Fig. 1b). In addition, the autoinhibited Apaf-1 is bound to ADP, but the CED-9-sequestered CED-4 is bound to ATP. Using the DALI server²³, the structures that are most homologous to CED-4 were found to be those of Cdc6 (ref. 24) and ORC2 (ref. 25), both ATPases belonging to the AAA+ (for ATPases associated with various cellular activities) family²⁶. This analysis, in conjunction with an earlier report²⁷, identifies CED-4 as a member of the AAA+ family of ATPases.

Mechanism of CED-4 recognition by CED-9

The structure of the CED-4–CED-9 complex reveals the molecular basis by which CED-9 recognizes CED-4 (Fig. 2a, b). The recognition interface of CED-9 primarily involves the N-terminal segment

(residues 67–79), the intervening loop between helices $\alpha 3$ and $\alpha 4$ (residues 143–147), and the $\alpha 6$ helix (residues 201–215). The N-terminal segment of CED-9 forms an extended loop, and stacks against the N-terminal portions of helices $\alpha 2$ and $\alpha 4$ and helix $\alpha 8$ of CED-4a (Fig. 2b). In this region, Asp 67 and Asp 79 in CED-9 each accept a pair of charge-stabilized hydrogen bonds from Arg 24 and Arg 117, respectively, in CED-4a. The $\alpha 6$ helix and the intervening loop between helices $\alpha 3$ and $\alpha 4$ in CED-9 interact with helix $\alpha 11$ and the following loop in CED-4a (Fig. 2b). Specifically, Arg 211 in CED-9 donates two hydrogen bonds to the backbone amide groups of Glu 214 and Asp 215 in CED-4a. Asn 212 in CED-9 also makes a hydrogen bond to Glu 52 in CED-4a. There are only a few van der Waals contacts at the CED-4a–CED-9 interface; these occur between the $\alpha 3$ – $\alpha 4$ intervening loop in CED-9 and helix $\alpha 11$ and the following loop in CED-4a (Fig. 2b).

The CED-4a–CED-9 interactions seem to be ideal for their functional regulation. The extensive interface, involving 3,550 Å² of buried surface area and 13 hydrogen bonds, provides exquisite recognition specificity between CED-4a and CED-9. Nonetheless, the shape complementarity at the interface is poor (Supplementary Fig. 2a). A large proportion of the buried amino acids at the interface are not involved in intermolecular interactions. These structural features make it easier for EGL-1 to release CED-4 from the CED-4–CED-9 complex.

Mechanism of CED-4 release by EGL-1

The structure of CED-9 in the CED-4–CED-9 complex is similar to that of isolated CED-9 (ref. 28), with an overall root mean square deviation (r.m.s.d.) of 0.7 Å, but it is quite different from that of CED-9 in the EGL-1–CED-9 complex²⁰. Comparison of the structure of the CED-4–CED-9 complex with that of the EGL-1–CED-9 complex²⁰ shows the molecular mechanism by which EGL-1 releases CED-4 from the inhibitory CED-4–CED-9 complex. The structural elements of CED-9 that directly bind to CED-4 are different from those that directly contact EGL-1 (Fig. 2c). However, the binding of EGL-1 to CED-9 is predicted to trigger pronounced conformational changes in CED-9 that result in the disruption of the interactions between CED-9 and CED-4a. On binding of EGL-1, the $\alpha 4$ helix of CED-9 is translocated towards CED-4a by as much as 6 Å (Supplementary Fig. 2b), which is predicted to cause a steric clash between CED-9 and CED-4a. In addition, four amino acids in CED-9 that are important for binding to CED-4 exhibit quite different side-chain and main-chain conformations upon binding to EGL-1 (Supplementary Fig. 2c). These altered conformations are no longer compatible for interaction with CED-4.

The binding affinity between CED-4 and CED-9 was estimated to be 48 ± 8 nM. In contrast, EGL-1 binds to CED-9 with a dissociation constant of approximately 6 nM (ref. 20). The stronger interaction between EGL-1 and CED-9 allows EGL-1 to bind to the CED-4-bound CED-9, and to dissociate CED-4 by inducing conformational changes in CED-9. To quantify this process directly, we titrated EGL-1 protein into pre-formed wild-type CED-4–CED-9 complex using isothermal titration calorimetry (ITC). EGL-1 efficiently dissociated CED-4 from the CED-4–CED-9 complex, with an apparent dissociation constant of 98 ± 18 nM (Fig. 2d).

CED-4a–CED-4b interface and ATP binding

One unexpected finding of this study is that a single CED-9 molecule binds to an asymmetric dimer of CED-4. The two molecules of CED-4 associate with each other mainly through van der Waals contacts (Fig. 3a, b). The asymmetric interaction results in the formation of an additional α -helix ($\alpha 11b$) in CED-4b that packs against CED-4a (Fig. 3b and Supplementary Fig. 1a). The interface between CED-4a and CED-4b consists of a primary and a secondary binding surface (Fig. 3a). In the primary interface, helices $\alpha 8$, $\alpha 11$ and $\alpha 11b$, all from the α/β -fold of CED-4b, pack against the CARD and α/β domains of

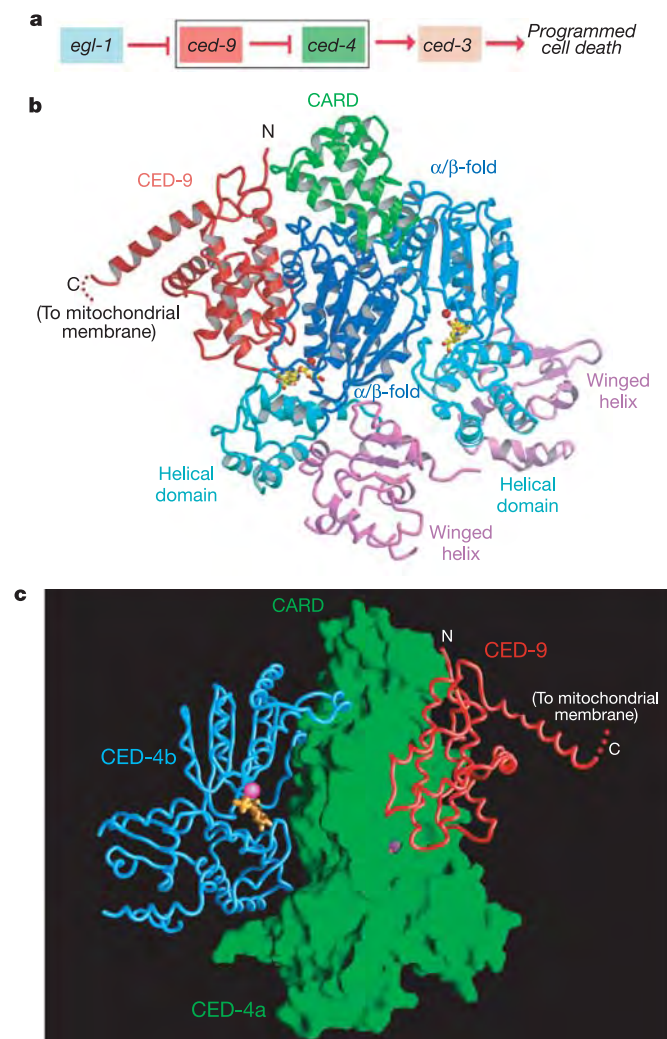


Figure 1 | Overall structure of the CED-4–CED-9 complex. **a**, A linear pathway of programmed cell death in *C. elegans*. **b**, Overall structure of the CED-4–CED-9 complex. Each CED-4 molecule comprises four sequential domains—CARD domain (green), α/β -fold (blue), helical domain (cyan) and winged-helix domain (magenta)—and contains a bound ATP molecule and a magnesium ion. **c**, Another view of the 2:1 CED-4–CED-9 complex. CED-4a (green) is shown as a surface representation. The bound ATP molecules and magnesium ions are highlighted in orange and magenta, respectively. Figures were prepared using MOLSCRIPT (ref. 39) and GRASP (ref. 40).

CED-4a (Fig. 3b). In the secondary area of contact, several residues from the helical domain of CED-4b make hydrogen bonds and salt bridges to the winged-helix domain of CED-4a.

Using mutagenesis, we tried to disrupt selectively the CED-4 dimer. Because of the extensive CED-4a–CED-4b interface, single point mutations only weakened, and failed to disrupt, their interactions (Supplementary Table 2). A CED-4 triple mutant, containing V230D, R233E and M234E amino-acid substitutions in helix α 12 (Fig. 3b), abolished formation of the CED-4 dimer and, consequently, formed an apparent 1:1 complex with CED-9 as determined by gel filtration (Supplementary Fig. 3a).

CED-4a and CED-4b have a similar structure, with an r.m.s.d. of 0.8 Å over 359 aligned C α atoms (Supplementary Fig. 3b). However, the CARD domain of CED-4a adopts a rigid conformation through interactions with the α / β -fold of CED-4b and CED-9, whereas the CARD domain of CED-4b is disordered in the crystals. Modelling studies show a severe steric clash if the CARD domain of CED-4b were to assume the same conformation as that of CED-4a. This analysis explains why CED-4b does not bind to CED-9: the CARD domain of CED-4b cannot adopt the same conformation that is required for interacting with CED-9, and the α / β domain of CED-4b

already uses a similar interface, which would be required for binding to CED-9, for interacting with CED-4a.

ATP and magnesium are largely buried in both CED-4 molecules (Fig. 3c). ATP binds along the hinge region between the α / β domain and the helical domain. The specific coordination of ATP is achieved by a total of eleven hydrogen bonds, ten of which are directed towards the phosphate groups (Fig. 3d). Only one specific hydrogen bond is made to the adenine base, with the N6 atom of ATP donating a hydrogen bond to the backbone amide of Tyr 131. It is important to note that this hydrogen bond dictates the specificity for adenine over guanine, as the O6 atom of guanine is unable to interact with an amide oxygen atom. The γ -phosphate group is coordinated by three hydrogen bonds from Lys 165 (in the P-loop), Arg 273 and Tyr 369 (Fig. 3d). The magnesium ion is bound by the β - and γ -phosphates, and by the side chain of Ser 166 in the P-loop.

Notably, the CED-4–CED-9 complex showed no detectable ATPase activity. This is probably due to the structural arrangement in the vicinity of the γ -phosphate, which does not allow a general base to activate a water molecule. The closest general base, Asp 251 in the Walker B box, accepts a pair of charge-stabilized hydrogen bonds from Arg 380 (Fig. 3d). The closest distance between the carboxylate

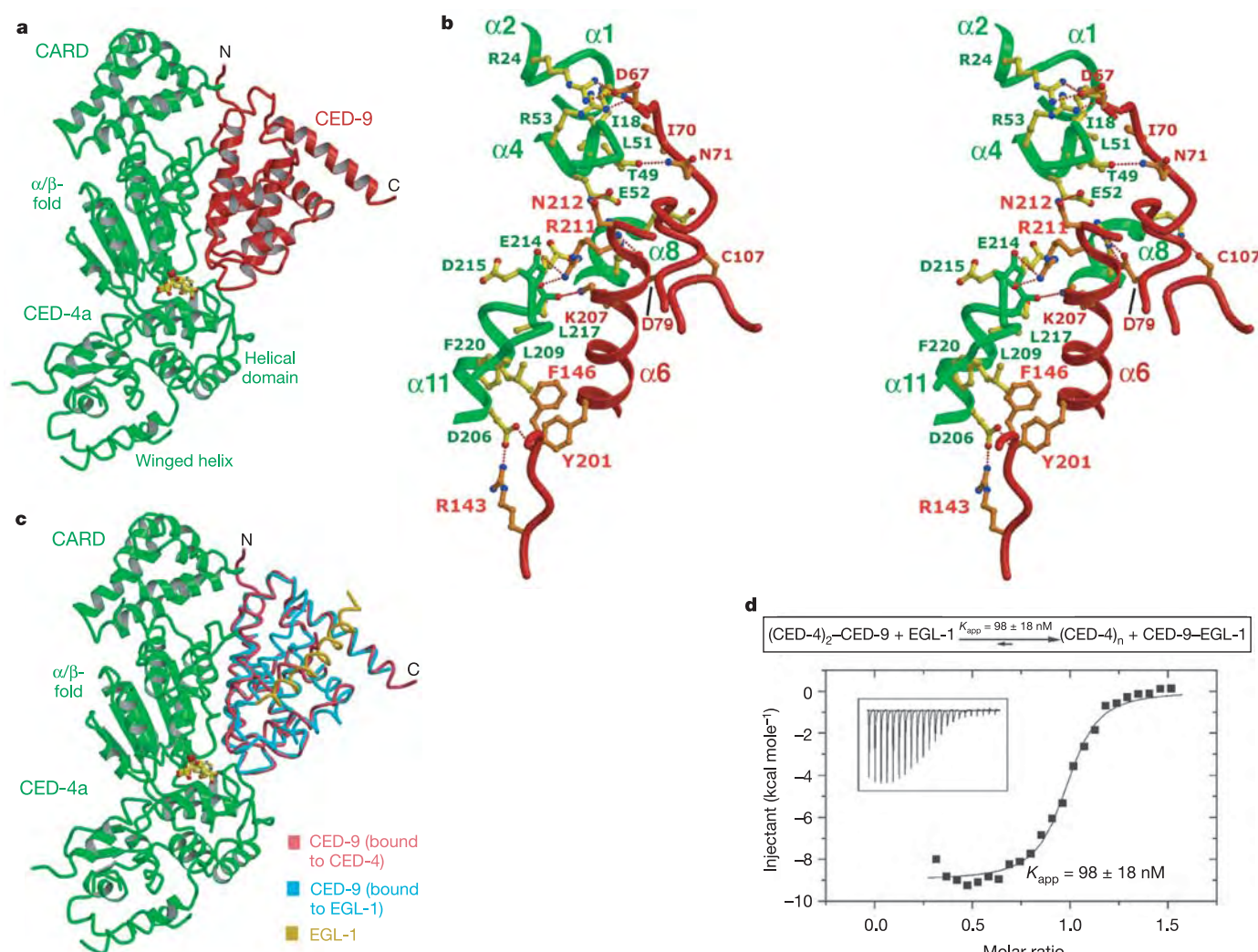


Figure 2 | Mechanisms of CED-4 recognition by CED-9, and CED-4 release by EGL-1. **a**, CED-9 binds to the CARD domain and the α / β -fold of CED-4a. **b**, A stereo view of the specific recognition between CED-4a (green) and CED-9 (red). Hydrogen bonds are shown as red dashed lines. The side chains from CED-4a and CED-9 are shown in yellow and orange, respectively.

c, Structural superposition of CED-4-bound CED-9 with EGL-1-bound CED-9. EGL-1 binding triggers a conformation change in CED-9 that destabilizes the CED-4–CED-9 interactions. **d**, EGL-1 releases CED-4 from the CED-4–CED-9 complex. An apparent dissociation constant of $98 \pm 18 \text{ nM}$ was derived from ITC experiments.

side chain of Asp 251 and the γ -phosphorous atom is greater than 6 Å.

Tetramerization of CED-4

EGL-1 binding to CED-9 results in the release of CED-4 from the inhibitory CED-4–CED-9 complex. It is unclear whether the released CED-4 dimer directly facilitates CED-3 activation, or whether it requires an additional step of oligomerization. To address this, we incubated the wild-type CED-4–CED-9 complex with excess EGL-1 and performed gel filtration analysis. CED-4 was eluted at a volume corresponding to a molecular weight of approximately 250 kDa (Fig. 4a, red line), about twice that of the CED-4 dimer. This result suggested that the released CED-4 dimer may undergo further dimerization.

To characterize the putative CED-4 tetramer, we examined its morphology by electron microscopy (Fig. 4b). About 9,500 raw individual particle images were manually, but non-discriminatorily, selected from electron micrographs. These particles were subjected to reference-free classification into 188 classes through a multivariate statistical analysis. The use of a large class number ensured that particles of different views were not assigned into the same class. Particles within the same class were averaged to provide statistically defined views. Three averages, related by an in-plane rotation, showed an apparent four-fold symmetry with a dimension of

13 nm (Fig. 4b). All other class averages, as anticipated, represented views of the tetramer with various degrees of out-of-plane tilt. No class average with three-, five- or six-fold symmetry was observed. These observations provide strong evidence that the released CED-4 dimer undergoes an additional step of dimerization.

Reconstitution of the CED-3 activation pathway

Despite strong genetic evidence, the cell death pathway from EGL-1 to CED-3 (Fig. 1a) has not been reconstituted *in vitro* using homogeneous recombinant proteins. We determined whether CED-4, CED-9 and EGL-1 can collectively control the activation of the CED-3 zymogen *in vitro*. We purified full-length wild-type CED-4, transmembrane-segment-deleted CED-9 (residues 1–251 and 48–251), and full-length (residues 1–91) and truncated EGL-1 (residues 31–87) (Fig. 5a). We also purified a stoichiometric CED-4–CED-9 complex and a monomeric CED-4 mutant (Fig. 5a).

The wild-type CED-3 zymogen, generated by *in vitro* translation, was slowly autoprocessed as previously reported^{9,29} (Fig. 5b, lanes 1–3). The autoprocessing requires the catalytic residue Cys 358, as the mutation C358S abolished CED-3 autoactivation (Fig. 5b, lanes 4 and 5). This result indicates that the processing of the CED-3 zymogen is not due to other contaminating protease(s). We first examined whether the recombinant full-length CED-4 protein can

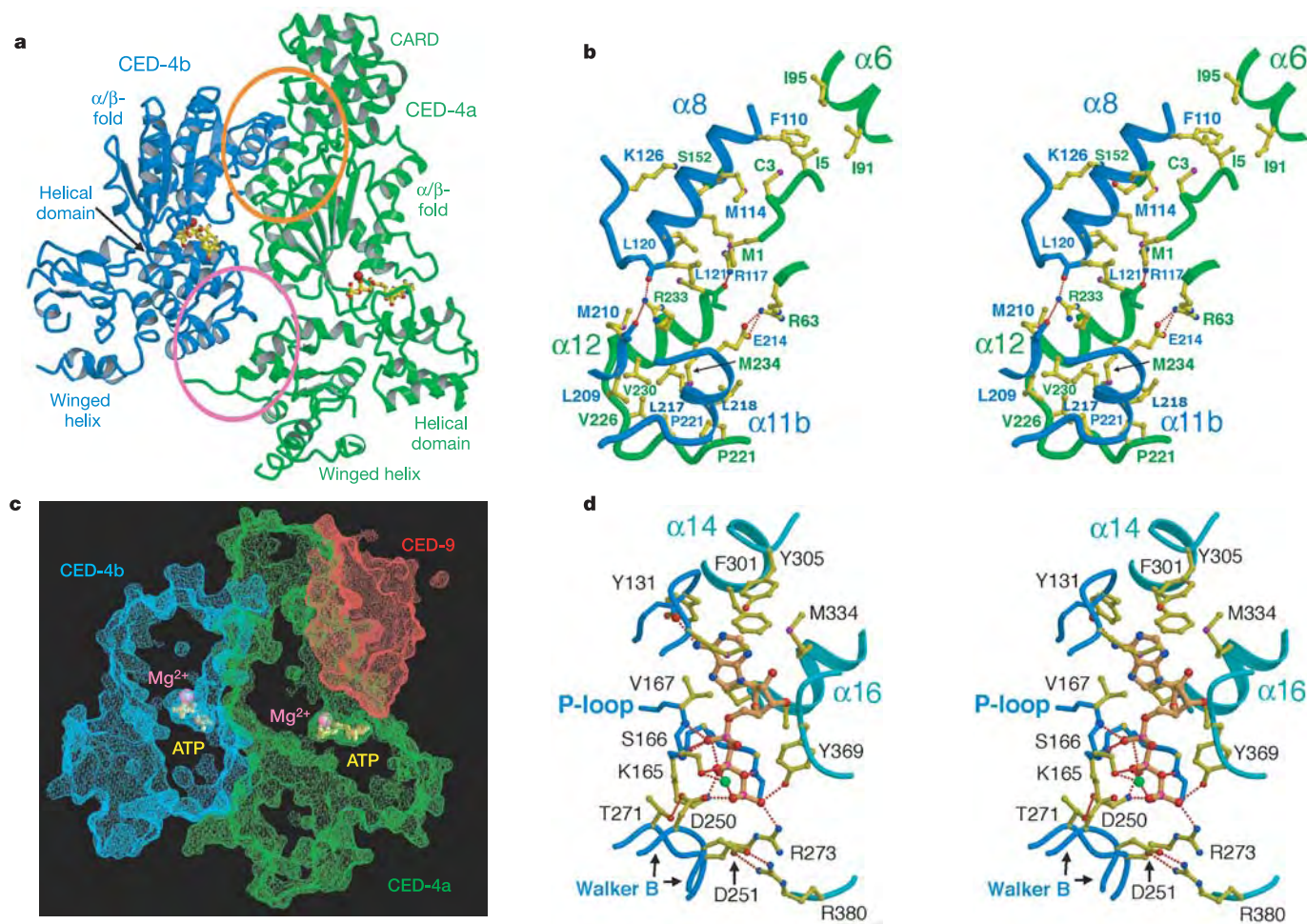


Figure 3 | Structural analyses of the CED-4a–CED-4b interface and ATP binding. **a**, The CED-4a–CED-4b interactions consist of a primary (orange circle) and a secondary (magenta circle) site. **b**, A stereo view of the primary interface of the CED-4 dimer. Helices $\alpha 8$, $\alpha 11$ and $\alpha 11b$, from CED-4b, stack against the CARD and α/β domains of CED-4a. **c**, ATP and magnesium are

buried in CED-4. ATP molecules and magnesium ions are highlighted in yellow and magenta, respectively. **d**, A stereo view of ATP coordination in CED-4. Hydrogen bonds are represented by red dashed lines. ATP is shown in orange, with nitrogen and oxygen atoms shown in blue and red, respectively.

facilitate the autoactivation of CED-3 (Fig. 5c). Under conditions where there was no detectable autoprocessing of CED-3 (Fig. 5c, lane 1), the autoactivation of the CED-3 zymogen was accelerated progressively with increasing amounts of CED-4 (Fig. 5c, lanes 2–6). This observation suggests that CED-4 alone is sufficient to promote CED-3 autoactivation.

Next, we examined the role of CED-9 on CED-4-mediated CED-3 activation (Fig. 5d). Under conditions where CED-4 promoted CED-3 maturation (Fig. 5d, lane 1), increasing amounts of CED-9 led to a progressive reduction of CED-3 autoactivation (Fig. 5d, lanes 2–5). Excess CED-9 completely suppressed the positive effect of CED-4 on CED-3 maturation (Fig. 5d, lane 5), reducing the amount of CED-3 cleavage to basal levels (Fig. 5d, lane 6). This result indicates that the anti-apoptotic function of CED-9 is not limited to passively sequestering CED-4 to mitochondria, but also to actively safeguard against CED-4-mediated CED-3 autoactivation. CED-9 does so by trapping an asymmetric dimer of CED-4 and preventing its further dimerization into a tetramer. In further support of this conclusion, incubation of the CED-4 tetramer with CED-9 led to the dissociation of the CED-4 tetramer and formation of the 2:1 CED-4–CED-9 complex (data not shown).

Finally, we assessed whether EGL-1 can counter CED-9-mediated inhibition of CED-4 (Fig. 5e). Under conditions where a stoichiometric complex of CED-4–CED-9 was unable to facilitate CED-3 maturation (Fig. 5e, lanes 1 and 2), incubation with increasing amounts of EGL-1 resulted in a progressive enhancement of CED-3 autoprocessing (Fig. 5e, lanes 3–6) due to the release of CED-4 from the inhibitory CED-4–CED-9 complex. Together, these assays have reconstituted the linear CED-3 activation pathway (Fig. 1a) using the recombinant homogeneous proteins CED-4, CED-9 and EGL-1. Our observations show that the genetically defined pathway is biochemically complete. The observed effects are specific, because neither CED-9 nor EGL-1 alone exhibited any effect on CED-3 maturation (Fig. 5f, lanes 3 and 4).

The reconstituted CED-3 activation assay provides us with a useful system to understand the mechanisms of CED-3 activation. The

CED-4 mutant (CED-4*; V230D-R233E-M234E), which formed a 1:1 complex with CED-9 (Supplementary Fig. 3a; Fig. 4a, pink dashed line), is exclusively monomeric in solution (Fig. 4a, green line). This mutant was unable to facilitate CED-3 autoactivation (Fig. 5g, lane 3). These observations suggest that the interface of the asymmetric CED-4 dimer is essential for the formation of the CED-4 tetramer, and for the subsequent induction of CED-3 autoactivation.

The CED-4* mutant (V230D-R233E-M234E) forms a stable dimer with another CED-4 mutant (CED-4"; L209E-L217E-L218E) that contains mutations on a different surface patch (Fig. 4a, purple line). This mutant CED-4 dimer is unable to dimerize further, suggesting that these mutations affect residues that are directly involved in CED-4 tetramerization, and thus are unlikely to affect interaction with CED-3. This mutant CED-4 dimer is also unable to facilitate the autoactivation of CED-3 (Fig. 5g, lane 5). Taken together, our data further suggest that only the wild-type CED-4 tetramer, but not a CED-4 dimer, is capable of mediating CED-3 autoactivation. These observations provide a mechanistic explanation as to why CED-9-sequestered CED-4 is unable to facilitate CED-3 autoactivation.

To examine the effect of CED-4 mutation *in vivo*, we introduced transgenes that expressed various CED-4 mutants, under the control of an endogenous *ced-4* promoter, into *ced-1(e1735);ced-4(n1162)* animals and examined whether they could rescue the *ced-4* cell-death-defective phenotype. No cell corpses were observed in *ced-1(e1735);ced-4(n1162)* animals, as the *ced-4(n1162)* mutation prevents almost all cell deaths. The wild-type *ced-4* transgene was able to restore cell killing in these animals (Supplementary Table 2). In contrast, the monomeric CED-4* mutant (V230D-R233E-M234E) induced very few cell deaths, indicating a loss of *ced-4* rescue activity. Other CED-4 mutants, with compromised abilities to form tetramers *in vitro*, also showed reduced capacities to induce cell death.

Discussion

Our study provides mechanistic explanations to published observations on the interaction of CED-4 with CED-9. CED-4L, a splicing variant of CED-4, inhibits programmed cell death³⁰. Compared with wild-type CED-4, CED-4L contains an insertion of 24 amino acids after Lys 212 at the centre of the CED-4a–CED-4b interface, and hence is predicted to disrupt this interaction. Therefore, CED-4L is unlikely to form a tetramer, and thus would be unable to mediate CED-3 autoactivation. Moreover, CED-4L (as CED-4a) can still form a dimer with wild-type CED-4 (as CED-4b). This would constitute dominant-negative regulation of wild-type CED-4, because wild-type CED-4 would be sequestered as a dimer and unable to facilitate CED-3 autoactivation. Finally, the regions surrounding Lys 212 have an important role in binding to CED-9, explaining the observation that CED-4L binds only weakly to CED-9 (ref. 31).

Neither CED-9-bound CED-4 dimer nor free CED-4 tetramer showed any detectable ATPase activity (data not shown), suggesting that ATP and magnesium only play a structural role in these complexes. This notion is supported by mass spectroscopic analysis, which only identified ATP, but not ADP, in both CED-9-bound CED-4 and free tetrameric CED-4. Thus, in contrast to Apaf-1, neither CED-9-bound CED-4 nor EGL-1-released CED-4 is an ATPase, and ATP hydrolysis does not seem to play a part in CED-3 activation in our *in vitro* assays. Nonetheless, we cannot rule out the possibility that an additional cofactor may help CED-4 to hydrolyse ATP *in vivo*, which further regulates CED-4 activity.

In summary, our study provides important insights into the regulation of cell death activation in *C. elegans*. In healthy cells, the CED-4 dimer is sequestered by CED-9 at mitochondria, and is unable to facilitate the autoactivation of CED-3 (Fig. 5h). At the onset of cell death, EGL-1 binds to, and triggers a conformational change in, CED-9, resulting in the release of CED-4 from the inhibitory CED-4–CED-9 complex. The released CED-4 dimer

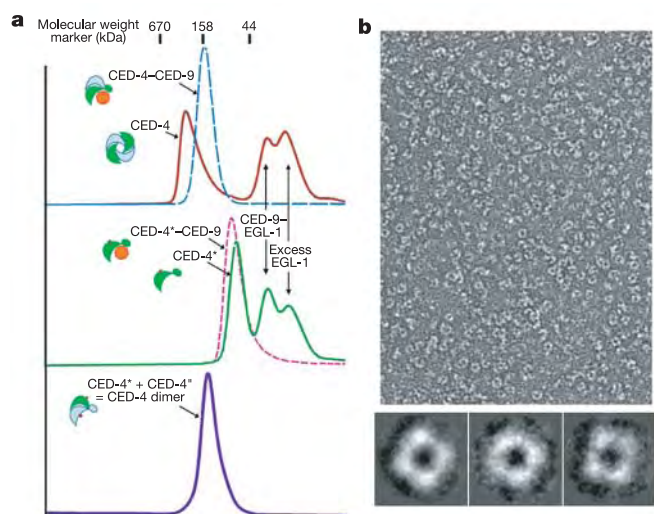


Figure 4 | Free CED-4 forms a tetramer. **a**, Free CED-4 appears to form a tetramer, as determined by gel filtration. Incubation of the CED-4–CED-9 complex (blue dashed line) with excess EGL-1 resulted in the release of CED-4 and formation of the EGL-1–CED-9 complex (red line). The mutant CED-4* (V230D-R233E-M234E) formed a 1:1 complex with CED-9 (pink dashed line) and is exclusively monomeric (green line). CED-4* and CED-4" (L209E-L217E-L218E) only formed a dimer (purple line). **b**, Electron microscopy reveals a tetrameric organization for free CED-4. A raw electron micrograph ($\times 30,000$) and three class averages of the CED-4 particle are shown.

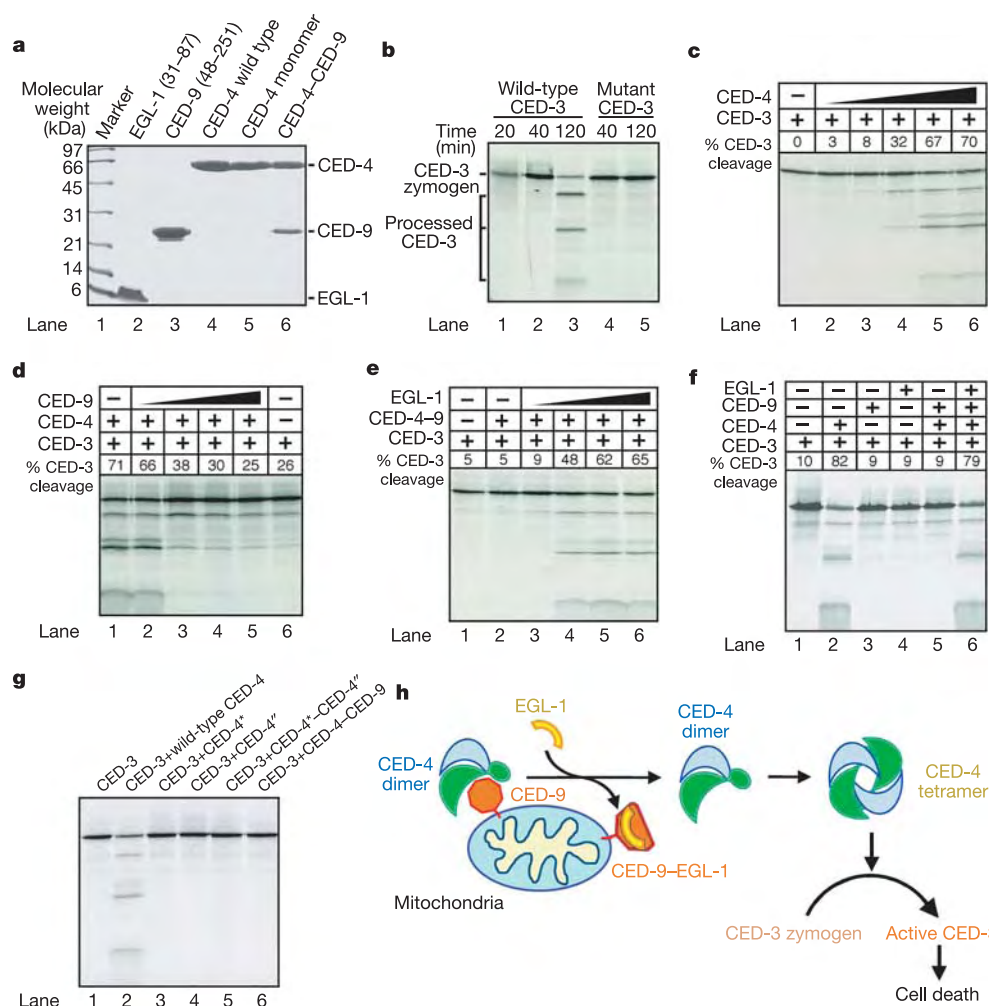


Figure 5 | Complete reconstitution of the CED-3 activation pathway using homogeneous proteins of CED-4, CED-9 and EGL-1. **a**, A representative SDS-PAGE gel (stained by Coomassie blue) showing the recombinant proteins. **b**, ^{35}S -labelled *in vitro* translated wild-type, but not C358S mutant, CED-3 zymogen was autoprocessed. **c**, Recombinant CED-4 protein facilitated the autoactivation of CED-3 in a concentration-dependent manner. Increasing amounts (0.1, 0.5, 2.5, 10 and 50 ng) of CED-4 (lanes 2–6) were used. **d**, Recombinant CED-9 protein inhibited CED-4-mediated autoactivation of CED-3 in a concentration-dependent manner. Increasing amounts (5, 25, 100 and 500 ng) of CED-9 (lanes 2–5) and 50 ng of CED-4

were used. **e**, Recombinant EGL-1 protein countered CED-9-mediated inhibition in a concentration-dependent manner. Increasing amounts (1, 5, 20 and 100 ng) of EGL-1 (lanes 3–6) and 50 ng of CED-4-CED-9 were used. **f**, Complete reconstitution of the CED-3 activation pathway, using 50 ng of each recombinant protein. **g**, Formation of a CED-4 tetramer is required for CED-3 autoactivation. Neither the monomeric CED-4* mutant nor the dimeric CED-4*-CED-4* mutant complex was able to facilitate the autoactivation of CED-3. Again, 50 ng of each recombinant protein was used. **h**, A schematic diagram of the cell death pathway leading to CED-3 activation.

undergoes further dimerization to form a tetrameric CED-4 complex, which is responsible for the induction of CED-3 autoactivation.

METHODS

Protein preparation. All constructs were generated using a standard polymerase chain reaction-based cloning strategy. CED-4 and CED-9 were coexpressed in the bacterial strain BL21 (DE3) using the vectors pBB75 and pET-15b (Novagen). CED-9 contained a 6 $\times$ His tag at its N terminus, and CED-4 was untagged. The bacterial growth temperature for soluble protein expression was optimized at 15°C. Proteins were purified to homogeneity as described²⁰.

Crystallization and data collection. Crystals of the CED-4-CED-9 complex were grown at 22°C using the hanging drop vapour diffusion method. The well buffer contained 0.1 M HEPES, pH 7.5, 20% (w/v) PEG 3350 and 0.2 M $(\text{NH}_4)_2\text{SO}_4$. The crystals grew to full size within 12 h, with a typical dimension of $0.3 \times 0.3 \times 0.8 \text{ mm}^3$. The crystals belong to the space group $P4_212$ and contain one complex per asymmetric unit. The unit cell has dimensions of $a = b = 128.9 \text{ \AA}$, $c = 209.9 \text{ \AA}$. Selenomethionine (SeMet)-labelled protein complex was crystallized under the same conditions. Native and anomalous diffraction data were collected at the National Synchrotron Light Source (NSLS)

beamline X-25. The data sets were collected at 100 K and processed using the HKL2000 software suite³².

Structure determination. The CED-4-CED-9 complex structure was solved by MAD phasing. At 5.0 Å resolution, 39 Se sites were found using the program SHELXD (ref. 33). Using the peak-wavelength data at 4.0 Å resolution, the correct handedness of the Se substructure was identified by the program ABS (ref. 34) and further refined using SOLVE (ref. 35), resulting in a figure of merit of 0.52 for all reflections within 3.1 Å. Combined with the native data, the 3.1 Å MAD phases were gradually extended to 2.6 Å by solvent flipping implemented in the program SOLOMON (ref. 36). The electron density map at 2.6 Å is of excellent quality and allows the model to be built in program O (ref. 37). The atomic model was refined using CNS (ref. 38).

Electron microscopy and image analysis. For electron microscopy, 5 μl of CED-4 sample, at a concentration of 0.1 mg ml^{-1} in 10 mM Tris, pH 8.0, 40 mM NaCl and 2 mM dithiothreitol (DTT), was transferred to a freshly glow-discharged 300-mesh copper grid covered with a thin layer of carbon film. The grid was then blotted and stained with a 5 μl drop of 2% uranyl acetate aqueous solution. The stain solution was blotted after 1 min, and the grid was left to air-dry. The grids were imaged in a JEOL 1200EX transmission electron microscope operated at 120 kV. Electron micrographs were recorded on Kodak SO-163 negative film at a magnification of $\times 30,000$. After digitizing the film

with a Nikon SuperCool scanner, ~9,500 particles were non-discriminatorily selected from the raw images. All particle images were classified through a multivariate statistical analysis. Particles belonging to the same classes were averaged to provide statistically better-defined views.

In vitro CED-3 activation assay. Wild-type full-length CED-3 was translated using the TNT T7 quick coupled transcription and translation system (Promega) at 30 °C for 25 min. The indicated proteins (CED-4 and its mutants, CED-9 and EGL-1) were added to the translation product in a 20 µl volume and incubated at 30 °C for another 25 min before samples were analysed by SDS–polyacrylamide gel electrophoresis (SDS–PAGE) and autoradiography. The extent of CED-3 cleavage, defined by the intensity ratio of cleaved CED-3 over total CED-3, was quantified using the EagleEye system (Stratagene).

Gel filtration assays. For each assay, 500 µl of purified wild-type or mutant CED-4–CED-9 complex, with or without EGL-1, was applied to Superdex 200 (Amersham) in buffer containing 25 mM HEPES, pH 8.0, 150 mM NaCl and 2 mM DTT. The peak fractions were analysed by SDS–PAGE and visualized by Coomassie blue staining.

Isothermal titration calorimetry. To obtain a direct binding affinity between CED-9 and CED-4, 0.1 mM CED-9 (residues 48–251) was titrated against 9 µM CED-4 monomer mutant using a VP-ITC microcalorimeter (MicroCal). To assess the ability of EGL-1 to dissociate CED-9 from the wild-type CED-4–CED-9 complex, 0.26 mM EGL-1 (residues 31–87) was titrated against 14 µM wild-type CED-4–CED-9 complex. All proteins were prepared in a buffer containing 25 mM HEPES, pH 8.0, and 150 mM NaCl. The data were fitted using the software Origin 7.0 (MicroCal).

ced-4 rescue assays. *ced-1(e1735);ced-4(n1162);unc-76(e911)* animals were injected with a *ced-4* construct at 50 ng µl⁻¹, along with the injection markers pTG96 (*Psur-5::GFP*) and p7616B (*unc-76* rescue plasmid) each at 20 ng µl⁻¹. Cell corpses were scored in the anterior head region of 1.5-fold and 4-fold transgenic fluorescent embryos. At least ten animals were scored for each transgenic line.

Received 7 April; accepted 4 July 2005.

- Horvitz, H. R. Worms, Life, and Death (Nobel Lecture). *ChemBioChem* **4**, 697–711 (2003).
- Horvitz, H. R. Genetic control of programmed cell death in the nematode *Caenorhabditis elegans*. *Cancer Res.* **59**, 1701–1706 (1999).
- Yuan, J., Shaham, S., Ledoux, S., Ellis, H. M. & Horvitz, H. R. The *C. elegans* cell death gene *ced-3* encodes a protein similar to mammalian interleukin-1β-converting enzyme. *Cell* **75**, 641–652 (1993).
- Xue, D., Shaham, S. & Horvitz, H. R. The *Caenorhabditis elegans* cell-death protein CED-3 is a cysteine protease with substrate specificities similar to those of the human CPP32 protease. *Genes Dev.* **10**, 1073–1083 (1996).
- Thornberry, N. A. & Lazebnik, Y. Caspases: Enemies within. *Science* **281**, 1312–1316 (1998).
- Yuan, J. & Horvitz, H. R. The *Caenorhabditis elegans* cell death gene *ced-4* encodes a novel protein and is expressed during the period of extensive programmed cell death. *Development* **116**, 309–320 (1992).
- Chinnaiyan, A. M., O'Rourke, K., Lane, B. R. & Dixit, V. M. Interaction of CED-4 with CED-3 and CED-9: a molecular framework for cell death. *Science* **275**, 1122–1126 (1997).
- Irmler, M., Hofmann, K., Vaux, D. & Tschopp, J. Direct physical interaction between the *Caenorhabditis elegans* 'death proteins' CED-3 and CED-4. *FEBS Lett.* **406**, 189–190 (1997).
- Seshagiri, S. & Miller, L. K. *Caenorhabditis elegans* CED-4 stimulates CED-3 processing and CED-3-induced apoptosis. *Curr. Biol.* **7**, 455–460 (1997).
- Wu, D., Wallen, H. D. & Nunez, G. Interaction and regulation of subcellular localization of CED-4 by CED-9. *Science* **275**, 1126–1129 (1997).
- Yang, X., Chang, H. Y. & Baltimore, D. Essential role of CED-4 oligomerization in CED-3 activation and apoptosis. *Science* **281**, 1355–1357 (1998).
- Hengartner, M. O. & Horvitz, H. R. *C. elegans* cell survival gene *ced-9* encodes a functional homolog of the mammalian proto-oncogene *bcl-2*. *Cell* **76**, 665–676 (1994).
- Chen, F. *et al.* Translocation of *C. elegans* CED-4 to nuclear membranes during programmed cell death. *Science* **287**, 1485–1489 (2000).
- James, C., Gschmeissner, S., Fraser, A. & Evan, G. I. CED-4 induces chromatin condensation in *Schizosaccharomyces pombe* and is inhibited by direct physical association with CED-9. *Curr. Biol.* **7**, 246–252 (1997).
- Spector, M. S., Desnoyers, S., Hoepfner, D. J. & Hengartner, M. O. Interaction between the *C. elegans* cell-death regulators CED-9 and CED-4. *Nature* **385**, 653–656 (1997).
- Conradt, B. & Horvitz, H. R. The *C. elegans* protein EGL-1 is required for programmed cell death and interacts with the Bcl-2-like protein CED-9. *Cell* **93**, 519–529 (1998).
- del Peso, L., Gonzalez, V. M. & Nunez, G. *Caenorhabditis elegans* EGL-1 disrupts the interaction of CED-9 with CED-4 and promotes CED-3 activation. *J. Biol. Chem.* **273**, 33495–33500 (1998).
- del Peso, L., Gonzalez, V. M., Inohara, N., Ellis, R. E. & Nunez, G. Disruption of the CED-9·CED-4 complex by EGL-1 is a critical step for programmed cell death in *Caenorhabditis elegans*. *J. Biol. Chem.* **275**, 27205–27211 (2000).
- Parrish, J., Metters, H., Chen, L. & Xue, D. Demonstration of the *in vivo* interaction of key cell death regulators by structure-based design of second-site suppressors. *Proc. Natl Acad. Sci. USA* **97**, 11916–11921 (2000).
- Yan, N. *et al.* Structural, biochemical, and functional analyses of CED-9 recognition by the proapoptotic proteins EGL-1 and CED-4. *Mol. Cell* **15**, 999–1006 (2004).
- Riedl, S. J., Li, W., Chao, Y., Schwarzenbacher, R. & Shi, Y. Structure of the apoptotic protease activating factor 1 bound to ADP. *Nature* **434**, 926–933 (2005).
- Zou, H., Henzel, W. J., Liu, X., Lutschg, A. & Wang, X. Apaf-1, a human protein homologous to *C. elegans* CED-4, participates in cytochrome c-dependent activation of caspase-3. *Cell* **90**, 405–413 (1997).
- Holm, L. & Sander, C. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138 (1993).
- Liu, J. *et al.* Structure and function of Cdc6/Cdc18: implications for origin recognition and checkpoint control. *Mol. Cell* **6**, 637–648 (2000).
- Singleton, M. R. *et al.* Conformational changes induced by nucleotide binding in Cdc6/ORC from *Aeropyrum pernix*. *J. Mol. Biol.* **343**, 547–557 (2004).
- Lupas, A. N. & Martin, J. AAA proteins. *Curr. Opin. Struct. Biol.* **12**, 746–753 (2002).
- Jaroszewski, L., Rychlewski, L., Reed, J. C. & Godzik, A. ATP-activated oligomerization as a mechanism for apoptosis regulation: fold and mechanism prediction for CED-4. *Proteins* **39**, 197–203 (2000).
- Woo, J. S. *et al.* Unique structural features of a BCL-2 family protein CED-9 and biophysical characterization of CED-9/EGL-1 interactions. *Cell Death Differ.* **10**, 1310–1325 (2003).
- Huginin, M., Quintal, L. J., Mankovich, J. A. & Ghayur, T. Protease activity of *in vitro* transcribed and translated *Caenorhabditis elegans* cell death gene (*ced-3*) product. *J. Biol. Chem.* **271**, 3517–3522 (1996).
- Shaham, S. & Horvitz, H. R. An alternatively spliced *C. elegans ced-4* RNA encodes a novel cell death inhibitor. *Cell* **86**, 201–208 (1996).
- Ottillie, S. *et al.* Mutational analysis of the interacting cell death regulators CED-9 and CED-4. *Cell Death Differ.* **4**, 526–533 (1997).
- Otwinski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Uson, I. & Sheldrick, G. M. Advance in direct methods for protein crystallography. *Curr. Opin. Struct. Biol.* **9**, 643–648 (1999).
- Hao, Q. ABS: A program to determine absolute configuration and evaluate anomalous scatter substructure. *J. Appl. Crystallogr.* **37**, 498–499 (2004).
- Terwilliger, T. C. & Berendzen, J. Automated structure solution for MIR and MAD. *Acta Crystallogr. D* **55**, 849–861 (1999).
- Abrahams, J. P. & Leslie, A. G. Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr. D* **52**, 30–42 (1996).
- Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Brunger, A. T. *et al.* Crystallography and NMR System: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Kraulis, P. J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
- Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct. Funct. Genet.* **11**, 281–296 (1991).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank H. R. Horvitz for complementary DNA sequences of *egl-1*, *ced-9*, *ced-4* and *ced-3*, and for his encouragement. We thank V. Dixit and X. Yang for providing *ced-3* constructs, and M. Becker and A. Saxena for beamtime at NSLS. This research was supported by NIH grants to Y.S. and D.X. H.L. acknowledges support from the Brookhaven National Laboratory LDRD program and a Department of Energy grant. The atomic coordinates of the CED-4–CED-9 complex have been deposited in the Protein Data Bank with the accession number 2A5Y.

Author Contributions N.Y. performed the bulk of the experiments. N.Y. and Y.S. designed and interpreted the bulk of the experiments. J.C., E.S.L., L.G., Q.L., J.H., J.-W.W. and D.K. contributed to experiments. J.C., E.S.L., L.G., Q.L., J.H., J.-W.W., D.K., H.L., Q.H. and D.X. contributed to data analysis and interpretation. N.Y. and Y.S. wrote the paper. J.C. and E.S.L. contributed equally to this work.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to Y.S. (yshi@molbio.princeton.edu).

ARTICLES

Structural insight into antibiotic fosfomycin biosynthesis by a mononuclear iron enzyme

Luke J. Higgins¹, Feng Yan², Pinghua Liu^{1,2}, Hung-wen Liu² & Catherine L. Drennan¹

The biosynthetic pathway of the clinically important antibiotic fosfomycin uses enzymes that catalyse reactions without precedent in biology. Among these is hydroxypropylphosphonic acid epoxidase, which represents a new subfamily of non-haem mononuclear iron enzymes. Here we present six X-ray structures of this enzyme: the apoenzyme at 2.0 Å resolution; a native Fe(II)-bound form at 2.4 Å resolution; a tris(hydroxymethyl)aminomethane-Co(II)-enzyme complex structure at 1.8 Å resolution; a substrate-Co(II)-enzyme complex structure at 2.5 Å resolution; and two substrate-Fe(II)-enzyme complexes at 2.1 and 2.3 Å resolution. These structural data lead us to suggest how this enzyme is able to recognize and respond to its substrate with a conformational change that protects the radical-based intermediates formed during catalysis. Comparisons with other family members suggest why substrate binding is able to prime iron for dioxygen binding in the absence of α -ketoglutarate (a co-substrate required by many mononuclear iron enzymes), and how the unique epoxidation reaction of hydroxypropylphosphonic acid epoxidase may occur.

Iron-dependent (S)-2-hydroxypropylphosphonic acid epoxidase (HppE) catalyses the final step in the biosynthesis of fosfomycin¹. Fosfomycin, or (1R,2S)-1,2-epoxypropylphosphonic acid, is an antibiotic used in the treatment of lower urinary tract infections², and is effective against methicillin-resistant³ and vancomycin-resistant⁴ strains of *Staphylococcus aureus*. In addition to the attention it receives as an antibiotic, fosfomycin is one of a few known natural products containing a carbon–phosphorus (C–P) bond. The biosynthesis of this unusual molecule in *Streptomyces wedmorensis* mirrors that of other C–P-bond-containing natural products with respect to the first two steps: the conversion of phosphoenolpyruvate (PEP) to phosphonopyruvate (PnPy) and the subsequent decarboxylation of PnPy to phosphonoacetaldehyde (PnAA)¹ (Fig. 1). After this, the pathways diverge, and PnAA is methylated by PnAA methylase in what is predicted by sequence and cofactor requirements to be a novel methylation reaction⁵. The product of this methylation, (S)-2-hydroxypropylphosphonic acid (S-HPP), is converted into the epoxide fosfomycin by HppE via an oxidative cyclization reaction with retention of the substrate hydroxyl oxygen atom^{6,7}. This latter transformation may be unique in biology⁷, as epoxidation reactions such as those catalysed by the P450 cytochrome family and non-haem iron dependent monooxygenases typically involve the conversion of substrate to an epoxide with oxygen atom insertion from dioxygen^{8,9}.

On the basis of sequence analysis, HppE is a member of the cupin superfamily, characterized by an antiparallel β -barrel (*cupa* is the Latin term for small barrel)¹⁰. Non-haem mononuclear iron-dependent proteins such as HppE are an interesting subset of this superfamily. These enzymes use a facial triad with two histidine ligands and one aspartic acid or glutamic acid, His₂(Glu/Asp), to catalyse a variety of different reactions, including DNA repair and antibiotic biosynthesis. An interesting question is how enzymes with

the same structure, metal and ligand triad can catalyse a diverse range of reactions.

With respect to the enzyme mechanism of HppE, biochemical experiments show that the reaction is initiated by a stereospecific and regiospecific hydrogen atom abstraction to yield a substrate radical intermediate en route to epoxide formation¹¹. A similar ring-closure mechanism has been proposed in the biosynthesis of penicillin by isopenicillin N synthase (IPNS)¹². Specifically, the bicyclic ring system of penicillin is derived from the four-electron oxidation of δ -(L- α -amino acidipoyl)-L-cysteinyl-D-valine (ACV). HppE and IPNS differ from most characterized cupin mononuclear iron enzymes in

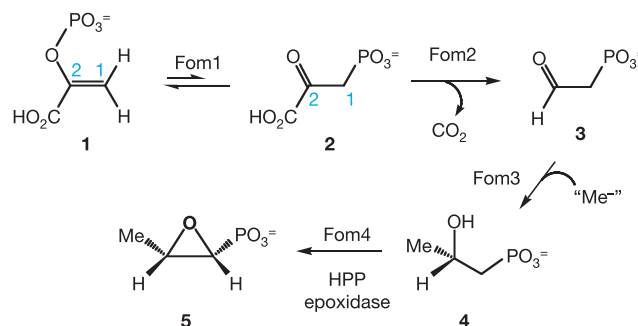


Figure 1 | Fosfomycin biosynthesis. The fosfomycin biosynthetic pathway requires phosphoenolpyruvate mutase (Fom1), phosphonopyruvate decarboxylase (Fom2), phosphonoacetaldehyde methyltransferase (Fom3) and HppE (Fom4). 1, phosphoenolpyruvate (PEP); 2, phosphonopyruvate (PnPy); 3, phosphonoacetaldehyde (PnAA); 4, (S)-2-hydroxypropylphosphonic acid (S-HPP); 5, fosfomycin. C1 and C2 positions are shown in blue.

¹Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ²Department of Chemistry and Biochemistry, University of Texas at Austin, Austin, Texas 78712, USA.

that these enzymes do not catalyse oxygen insertion into substrate or co-substrate. Also, both HppE and IPNS are independent of cofactors/co-substrates such as α -ketoglutarate (α -KG), tetrahydropterin, ascorbate and iron-sulphur clusters, used by other mononuclear non-haem iron proteins as a source of reducing equivalents⁷. The reason for their independence from cofactors is different, however. Unlike fosfomycin biosynthesis, the conversion of ACV to penicillin *N* is a four-electron oxidation coupled to a four-electron reduction of dioxygen to water. Therefore, IPNS does not require an external electron source. In contrast, HppE requires two electrons transferred from an endogenous reductase, in addition to the two electrons supplied by the substrate, to completely reduce dioxygen to water. The putative HppE reductase remains to be identified.

Here we present the first crystal structures of HppE, a representative of a new subfamily of mononuclear iron enzymes, and discuss how this enzyme is specifically tuned to facilitate the unusual radical-mediated epoxidation reaction found in the pathway of fosfomycin biosynthesis.

Overall structure of HppE

The three-dimensional structure of HppE (Supplementary Tables 1 and 2) confirms its identity as a member of the cupin superfamily, possessing the requisite β -barrel fold in which antiparallel β -strands are wrapped around a barrel core in a jelly roll variant of a Greek key or β -sandwich motif (Fig. 2a). As expected, the epoxidase is a physiological homotetramer with one iron per monomer¹. Each monomer consists of two domains: the α -domain, which is all α -helical, and a β -domain, which consists of anti-parallel β -strands in a jellyroll β -barrel motif (Fig. 2a). As predicted by sequence alignments¹³, the facial triad—His 138, Glu 142 and His 180—is housed within the β -barrel, defining the HppE active site. The spacing of the first two triad members (His 138 and Glu 142) is different in HppE (HxxxE) and germin oxalate oxidase (HxxxxE)

relative to the majority of other proteins within the superfamily (HxD/E), an important consideration for bioinformatics studies. The smaller α -domain has only 47 residues, consisting of five helices, and represents a unique helical domain fold (Fig. 2a) according to a Dali server database search. An extended 32-residue linker separates the α - and β -domains of the HppE such that the closest distance between the two domains is approximately 15 Å (Fig. 2a). This separation translates to a striking quaternary structure that is unique in the cupin superfamily¹⁴ (Fig. 2b, c).

The HppE active site

Structural comparison of apo-HppE with Fe(II)-HppE shows that regions around the metal binding site in the β -domain, which are disordered in the apo-structure, become ordered upon metal binding. The ordering of strand 5 and the loop between strands 5 and 6 brings triad ligand His 138 into proximity of the iron atom (Fig. 3a, b). In one of three molecules in the asymmetric unit of the Fe(II)-HppE structure, two waters occupy exchangeable sites yielding a five-coordinate metal centre (Fig. 3b). In the other two molecules, the electron density is less well defined but is consistent with direct coordination of three water molecules to Fe(II) (data not shown).

In the mononuclear iron proteins, residues responsible for substrate specificity, regioselectivity and stereoselectivity are found above the ligand triad on the other end of the β -barrel¹⁵. On the basis of the overall HppE structure, residues on the 'tunable' face of the β -domain that are potentially significant for catalysis include Tyr 105, Tyr 103, Tyr 102 and Arg 97. In HppE, both β - and α -domains contribute to the active site. Helices 1 and 2 of the α -domain comprise one side of the active site (Fig. 4b) and Lys 23 from the α -domain is positioned near the metal centre (Fig. 3). Interestingly, HppE is known to selectively oxidize Tyr 105 to dihydroxyphenylalanine (DOPA) via a self-hydroxylation reaction¹⁶, correctly predicting that Tyr 105 lies near the iron site (Fig. 3b). In a Y105F mutant protein, Tyr 103 is oxidized to DOPA¹⁶; this is

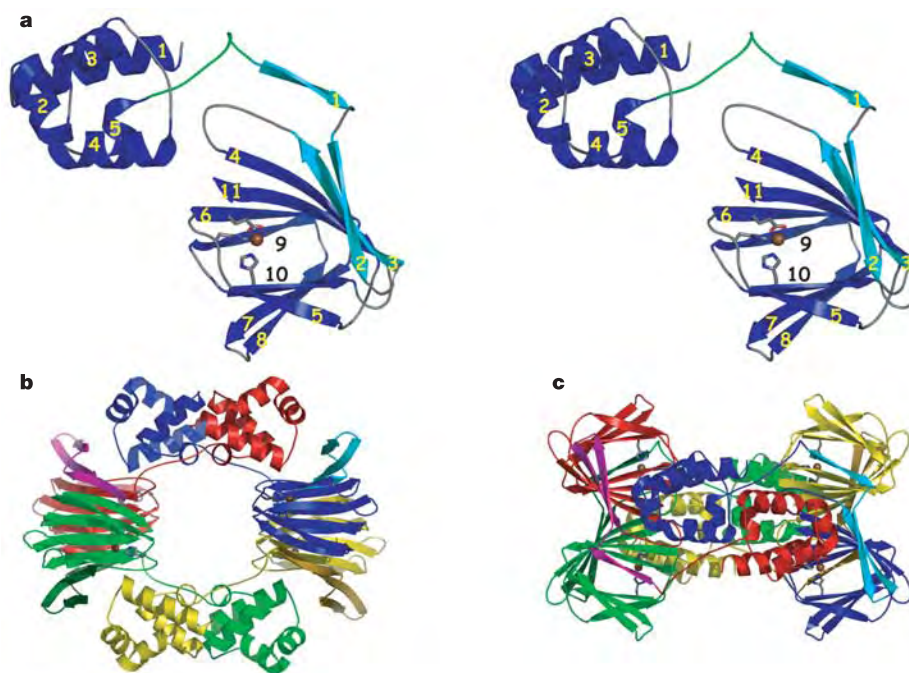


Figure 2 | Overall structure of Fe(II)-HppE. **a**, An HppE monomer consists of an α -domain (blue), an interdomain linker (green) to a single β -strand 1 (cyan) and a β -domain (blue). This stereoview highlights the cantilever hairpin (β -strands 2 and 3) in cyan, facial triad ligands (Glu 142, His 138 and His 180) in ball-and-stick, and iron as a brown sphere. Helices and strands are numbered separately and sequentially with respect to the primary

structure. **b**, HppE tetramer, coloured by molecule, is shown down one of the three two-fold axes of symmetry. The cantilever hairpin is coloured cyan in the blue molecule, magenta in the red molecule, dark yellow in the yellow molecule, and dark green in the green molecule. **c**, HppE tetramer oriented along a second two-fold axis of symmetry, orthogonal to that in **b**. The cantilever hairpins are coloured as in **b**. Figs 2–4 were made in PyMol³¹.

consistent with Tyr 103 being close to the iron site, but not as close as Tyr 105, as it is only modified in the mutant protein (Fig. 3b). Although HppE is not a monooxygenase with respect to the biosynthesis of fosfomycin, it does catalyse these self-hydroxylation reactions in the absence of substrate, a phenomenon observed in other family members such as taurine/ α -KG dioxygenase (TauD)¹⁷. Although self-hydroxylation is not part of the fosfomycin synthesis pathway, it is interesting in terms of the reactivity and the plasticity of the HppE active site. In the structure of Fe(II)-HppE, Tyr 105 is within 8.7 Å and Tyr 103 is within 10.8 Å, indicating that regions of the active site must be conformationally flexible to allow these tyrosine residues to approach the iron for hydroxylation to occur. Results presented below confirm this conformational flexibility.

Substrate binding and conformational changes

To investigate substrate binding without turnover, Co(II)-HppE was used as a catalytically inert model system⁷, and anaerobic crystals of Fe(II)-HppE were obtained in two different crystal forms (form-1, $P6_522$ and form-2, $P4_22_12$) and soaked with substrate (Supplementary Tables 1 and 2). The presence of a moderately heavy atom in the substrate (that is, phosphorous) gave rise to 6–9 σ omit electron density peaks indicating the position of the phosphonic acid moiety and the orientation of the substrate in these structures. Notably, these structures reveal two modes of substrate binding, suggesting a

two-step binding process before catalysis. In one of the two molecules present in the asymmetric unit of Co(II)-HppE and form-1 Fe(II)-HppE, the substrate binds in a monodentate fashion with an oxygen atom of the phosphonic acid moiety coordinating the metal and displacing one bound water (Fig. 3c). In the other molecule of Co(II)-HppE and form-1 Fe(II)-HppE, as well as in form-2 Fe(II)-HppE, S-HPP forms a bidentate interaction with the metal centre via the 2-hydroxyl oxygen and an oxygen atom from the phosphonic acid moiety, substituting for water ligands in the free form, and leaving a third coordination site open for subsequent dioxygen binding (Figs 3d and 4a). In both Fe(II) and Co(II) crystal forms, bidentate binding of S-HPP induces a conformational change in the β -barrel, whereas substrate bound in a monodentate fashion physically precludes this conformational change from occurring. The conformational change involves β -strands 2 and 3 (Fig. 2), which form a β -hairpin substructure and act as a cantilever, responding to the positioning of substrate in the bidentate orientation by moving in and covering one side of the active site (Fig. 4b). Movements near the hinge, or fulcrum, of the hairpin are subtle whereas those near the end are more appreciable (Fig. 4b). Specifically, Tyr 105 moves towards the metal centre by 1.8 Å, positioning its hydroxyl group within hydrogen-bonding distance of a phosphonate oxygen of S-HPP (2.7 Å), while Tyr 102 moves 8.2 Å towards the metal, occupying approximately the same position as Tyr 103 in the Fe(II)-HppE structure (Fig. 3b, c). In addition, Arg 97, on the amino terminus of strand 2, swings towards the metal centre through a distance of approximately 7.4 Å (Fig. 4b). There is little change in other proximal residues of the β -barrel. Relative to the cantilever fulcrum defined by residues Ile 92 and Asn 106, the β -hairpin swings through an angle of 29 degrees, decreasing the solvent-exposed surface area of the

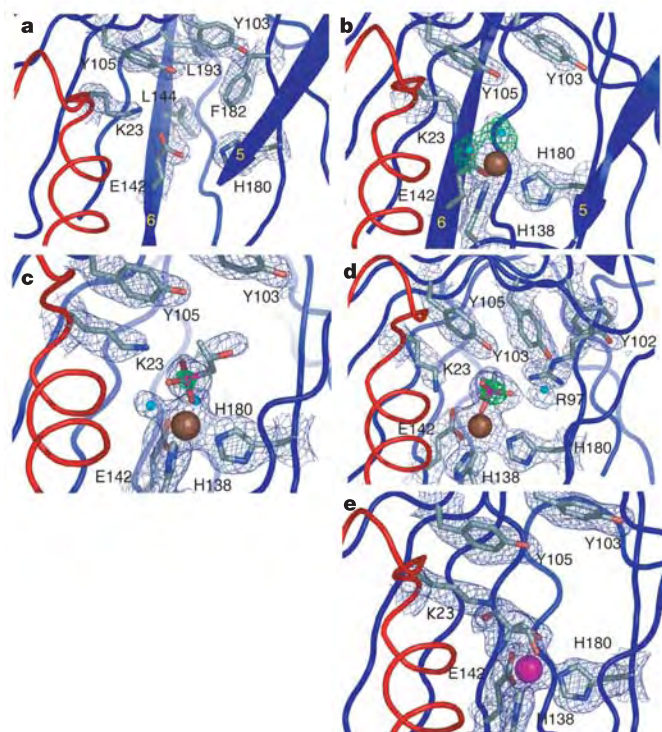


Figure 3 | HppE active sites displayed with $2F_o - F_c$ maps contoured from 1.0–1.5 σ . **a**, Apo-HppE active site with strands 5 and 6 labelled, and the α -domain (red), β -domain (blue) and carbon (grey), oxygen (red) and nitrogen (blue) positions shown. **b**, Fe(II)-HppE active site. $F_o - F_c$ electron density peaks contoured at 3 σ indicate the positions of water molecules (cyan) bound to Fe(II) (brown). **c**, Monodentate substrate binding in form-1 S-HPP-Fe(II)-HppE active site, coloured as in **a**, **b** with phosphorus in magenta. An omit map contoured at 6 σ shows an electron density peak (green) coincident with the phosphorus atom here and in **d**. **d**, Bidentate substrate binding in the form-2 S-HPP-Fe(II)-HppE active site with domains coloured as in **a** and atoms coloured as in **a**–**c**. **e**, The Tris-Co(II)-HppE active site, coloured as in **a** with cobalt shown as a magenta sphere. See Supplementary Fig. S1 for stereo images of S-HPP-Fe(II)-HppE and Supplementary Fig. S3 for S-HPP-Co(II)-HppE.

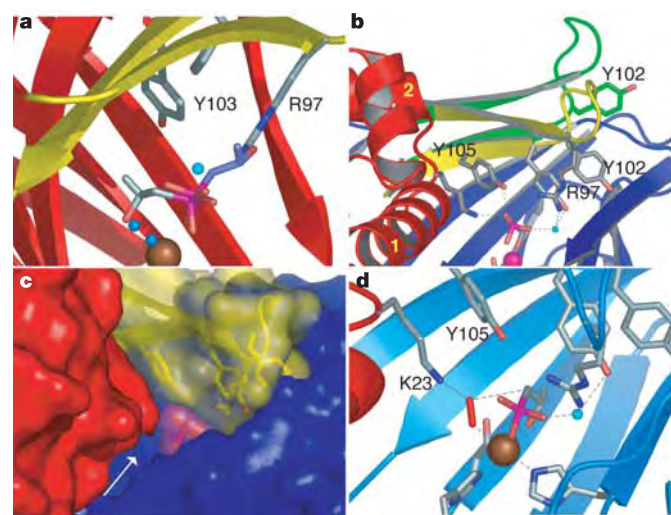


Figure 4 | Structural insights into catalysis. **a**, Superposition of monodentate and bidentate substrate binding modes in Fe(II)-HppE with the cantilever hairpin (yellow) in the closed conformation. Atoms coloured as in Fig. 3, with water molecules in the monodentate structure (blue) and bidentate structure (cyan). **b**, A β -hairpin changes conformation between substrate-free (ribbons and residues in green) and substrate-bound (ribbons in yellow, residues in grey) states of Fe(II)-HppE. Both positions of Tyr 102 are labelled. Ile 92 and Asn 106, which represent the start and end of the hairpin, respectively, are positioned under helix 2, and are not visible in this orientation. **c**, Van der Waals surface representation after hairpin closure, coloured as in Fig. 4b. White arrow indicates a proposed channel for dioxygen. **d**, Model for a dioxygen-bound structure. Dashed lines illustrate potential ligand-metal and dioxygen-side chain interactions, as well as a possible path for hydrogen atom abstraction from an iron-hydroperoxo intermediate (mechanism B in Fig. 6). See Supplementary Fig. S2 for stereo images.

β -barrel. In addition to the linear motion towards the active site, the right-handed twist (Fig. 4b) of the hairpin increases upon substrate binding, conserving the hydrogen-bonding interactions of the β -sheet. Alignment of the two substrate binding modes reveals that the monodentate binding position of S-HPP sterically precludes motion of strand 3 towards the metal centre, in that it occupies the position that Arg 97 holds in the closed conformation (Fig. 4a).

The movement of the cantilever does not entirely seal off the active site. Dioxygen can still access the iron, and the negatively charged phosphonic acid moiety of S-HPP is still exposed to solvent (Fig. 4c). In contrast, the hydrophobic portion of S-HPP, where hydrogen atom abstraction occurs, is completely buried in a small hydrophobic cavity. The conformational change also creates an extensive hydrogen-bonding network between the substrate, bound water molecules and the enzyme (Fig. 5). Mutation of two of the residues involved in the hydrogen-bonding network, Y105F and K23A, leads to complete loss of activity as determined by a ^{31}P -NMR assay, under conditions where wild-type HppE has activity and ligand triad mutant E142A does not. The movement of the cantilever also positions three Tyr residues (102, 103, 105) in a row, leading from the active site to the surface of the protein (Fig. 3d). This suggests that these Tyr residues could play a role in electron transfer between the putative reductase and the iron in the active site.

The structure of tris(hydroxymethyl) aminomethane (Tris) in complex with Co(II)-HppE provides insight into substrate specificity and active site closure (Fig. 3e). The Tris molecule is best fitted to the electron density when modelled as binding to the Co(II) metal centre via two hydroxymethyl oxygens and the nitrogen of the aminomethane moiety (Fig. 3e). This coordination geometry places the remaining hydroxymethyl substituent within hydrogen-bonding distance of Lys 23 (2.9 Å). With Tris in the active site, one strand of the hairpin cantilever is in the substrate-free position while the other strand is disordered. Given that the physiological substrate of HppE

is smaller than Tris, is charged instead of neutral, and has additional hydrogen bond acceptors, it is not surprising that the hairpin cantilever does not close in as it does when the bidentate substrate is bound. Thus, the cantilever hairpin is only positioned for catalysis when a molecule of the correct size is bound in the appropriate conformation.

Mechanism of HppE

Over the past ten years, there have been a series of structural studies of mononuclear iron proteins, including work on enzymes involved in antibiotic biosynthesis such as α -KG-independent IPNS^{12,18} and α -KG-dependent deacetoxycephalosporin C synthase (DAOCS)¹⁹. The structures of HppE described here, together with previous biochemical studies^{1,7,11,16}, allow us to compare HppE with these other enzymes to probe the fundamental properties of this enzyme subfamily. Mononuclear iron proteins in the cupin family must face a series of mechanistic challenges: (1) fine-tuning of the protein structure to bind substrate, co-substrate and/or cofactor, (2) activation of iron for dioxygen binding, (3) acquisition of reducing equivalents and (4) protection of radical intermediates during catalysis. With this structural information, we can consider how HppE accomplishes the above, and why it differs from so many other family members in not using α -KG as a co-substrate.

The mechanism of substrate binding in HppE has been investigated using biochemical and biophysical techniques⁷. As shown by electron paramagnetic resonance (EPR) spectroscopy, the metal environment is heterogeneous when substrate binds to Fe(III)-HppE. This heterogeneity coalesces into a single homogeneous species in the EPR spectrum of Fe(II)-HppE bound with substrate and nitric oxide, a catalytically inert surrogate for dioxygen⁷. These EPR data indicate that the substrate binds directly to iron, and are consistent with multiple substrate-binding modes. In the crystal structures, we find that the substrate binds directly to the metal and that there are two binding modes, monodentate and bidentate. As there are no lattice contacts to explain the multiple substrate conformations, and as the two conformations are found with both Co(II)- and Fe(II)-reconstituted enzymes, we propose that both binding modes are relevant, and that S-HPP binds in a two-step process. In this proposed two-step process, S-HPP first binds in a monodentate fashion via an oxygen atom of the phosphonic acid substituent such that a water molecule is displaced. Monodentate binding is followed by rotation about the C1–P bond and the C1–C2 bond of the substrate, allowing for bidentate coordination of S-HPP to the metal centre concomitant with the displacement of remaining water molecules (Figs 4b and 6). With the substrate in a bidentate configuration, the charge and spatial complementarity are such that they afford the 8 Å shift of the cantilever hairpin, covering the hydrophobic portion of the substrate bound in the active site.

Following substrate binding, the next steps in the mechanism are iron activation and dioxygen binding. Here, the negative charge of substrate could serve to enhance the affinity of dioxygen for the iron centre by favouring higher oxidation states, a common mechanistic theme for this family of proteins^{9,19–22}. It is interesting that HppE has evolved to recognize this unusual C–P bond-containing natural product, and to use its physical properties to activate iron for dioxygen binding. The active site iron is activated for dioxygen binding with the substrate configured in either monodentate or bidentate modes. In either conformation of the cantilever, dioxygen can access the iron through a very small channel created at the interface of the α - and β -domains (Fig. 4c). This small channel leads to the only open coordination site on the metal in the bidentate S-HPP complex. This coordination site is where we have modelled dioxygen (Fig. 4d).

After formation of a ternary complex with substrate and dioxygen, C1 hydrogen atom abstraction by an activated Fe-oxo species yields a transient substrate radical intermediate that undergoes cyclization to yield fosfomycin¹¹. There are several plausible mechanisms for active

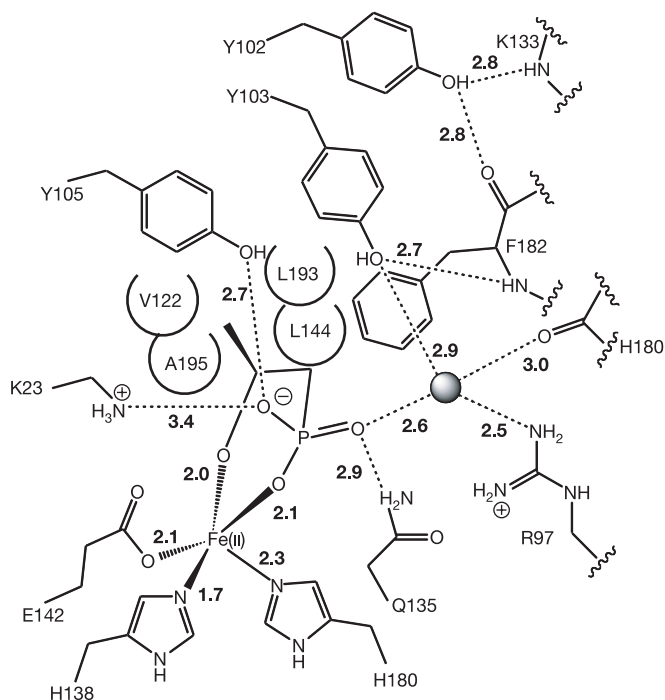


Figure 5 | Schematic of interactions of bidentate-bound S-HPP with protein, water and Fe(II). Water shown as a grey sphere. Residues within the substrate hydrophobic cavity are labelled (semicircles) and include Phe 182. Hydrogen-bonding contacts (dotted lines) are shown with distances indicated in Å. For simplicity, the interaction of the Arg 97 guanidinium with the phosphonic acid moiety (~3.4–3.6 Å) is not shown.

site reduction and epoxide ring formation, including hydrogen atom abstraction via a Fe(IV)-oxo intermediate (mechanism A) and hydrogen atom abstraction by a Fe(III)-hydroperoxide intermediate (mechanism B) (Fig. 6). The four-electron oxidation of ACV catalysed by IPNS uses both direct hydrogen atom abstraction by an Fe-superoxo species, a variation of mechanism B, and substrate radical formation via a Fe(IV)-oxo species, as in mechanism A^{12,23}. The hydrogen atom abstraction is both stereo- and regiospecific¹¹, and the dioxygen-binding model in Fig. 4d can explain the observed stereo- and regiospecificity. An oxygen species bound in the open coordination site could only access the relevant hydrogen on C1, the pro-*R* hydrogen. The C2 hydrogen points away from the oxygen-binding site, and is protected from abstraction by the C2 methyl substituent.

If the model in Fig. 4d is correct, then Lys 23 from the neighbouring molecule should have a role in catalysis as it is positioned directly above the proposed dioxygen-binding site. In fact, replacement of Lys 23 with Ala results in a protein with no detectable activity, supporting this model. The exact role of Lys 23 is unclear. In both mechanisms A and B, the negatively charged Fe-peroxide intermediate would need to be stabilized and subsequently protonated. Lys 23 could be involved in both processes.

Like other mononuclear iron enzymes, HppE and IPNS must shield reactive intermediates from solvent. Both enzymes have evolved different strategies to meet this need. In IPNS, ACV binding facilitates ligand exchange, triggering the extension of the carboxy-terminal helix by five residues, which seals the β -barrel¹². In HppE, the cantilever hairpin moves towards the metal centre concomitant with bidentate substrate binding, serving to partly occlude the β -barrel (Fig. 4b). Comparison of Figs 3c–e shows that the structure of HppE is finely tuned to respond to the correct substrate bound in an appropriate orientation (bidentate binding) before adopting this closed catalytic conformation.

Structural and mechanistic comparisons

Structural comparisons of HppE, IPNS and DAOCS suggest why HppE is an α -KG-independent enzyme, in contrast to the majority of cupin mononuclear iron proteins. First, one must consider that α -KG can serve two functions: it can serve as a source of two electrons for dioxygen reduction, and it can activate iron for dioxygen binding. With respect to the latter function, the bidentate binding of α -KG to iron via its negatively charged C1 carboxylate and the 2-oxo substituent primes the iron for dioxygen binding and subsequent iron oxidation²⁴. This function is not necessary in HppE as the substrate is negatively charged and can independently activate iron for dioxygen binding. Similarly, IPNS, another α -KG-independent enzyme within this subfamily, facilitates dioxygen binding via coordination of a negatively charged thiolate substituent of the substrate (ACV)¹². A second consideration is that bidentate substrate binding in HppE fills two of three open coordination sites on iron, preventing simultaneous bidentate binding of α -KG. Most characterized α -KG-dependent iron enzymes simultaneously bind α -KG and substrate in their catalytic positions, so that a highly reactive Fe-oxygen species is not formed until substrate is available⁹. A notable exception, DAOCS, binds α -KG and the substrate penicillin to the same coordination site on the iron¹⁹. In this case, DAOCS has evolved a mechanism by which the Fe-oxygen species formed by reaction with α -KG is 'booby-trapped' (that is, the reactivity is contained) until the trap is sprung by substrate binding¹⁹. Although HppE could have evolved a similar protection mechanism to allow ping-pong binding of α -KG and substrate to the iron, this enzyme does not require α -KG for activation, and thus a better solution appears to be the use of the putative HppE reductase. Although all cupin mononuclear iron proteins have to activate iron for dioxygen binding and obtain reducing equivalents for the conversion of dioxygen to water, the strategies used are different. When coupled to substrate oxidation, α -KG can fulfill both requirements. In IPNS,

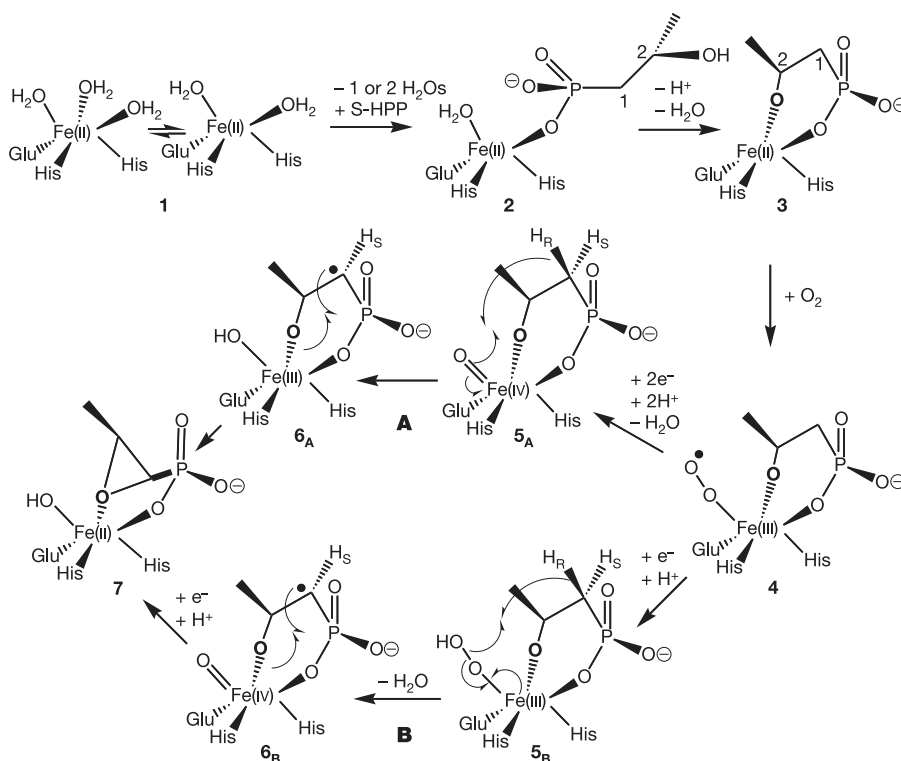


Figure 6 | Possible mechanisms for HppE. In a two-step binding model, the substrate binds iron first as a monodentate (2) and then a bidentate ligand (3), displacing 2–3 bound waters. Substrate binding activates the iron for dioxygen binding (4). The mechanisms diverge at intermediate (4). In

mechanism A, an Fe(IV)-oxo species abstracts the hydrogen atom (5_A). In mechanism B, the hydrogen atom is abstracted by an iron-hydroperoxo intermediate (5_B). For both mechanisms, the formation of a substrate radical (6) leads to ring closure and product formation (7).

the substrate alone can fulfill both roles. In HppE, the substrate facilitates dioxygen binding but cannot provide all of the electrons, and thus this enzyme requires an endogenous reductase. In the latter respect, HppE deviates from other cupin mononuclear iron enzymes.

This work presents the first structures of HppE, revealing a new subfamily of cofactor-independent mononuclear iron proteins. Structural analysis provides insight into the mechanism of substrate binding and catalytic initiation en route to the biosynthesis of fosfomycin. The growing threat of antibiotic-resistant microorganisms accentuates the importance of understanding the chemical mechanisms governing antibiotic biosynthesis. This knowledge can be applied to protein engineering of mononuclear iron scaffolds and the biosynthesis of unique antibiotic analogues.

METHODS

Site-directed mutagenesis, protein purification, assays of mutant and wild-type proteins, crystallization and data collection were carried out as described (Supplementary Methods). To determine the X-ray structures of HppE, experimental phases were obtained using data from the Tris-Co(II)-SeMet-HppE crystals. A single data set, collected at the Se absorption peak wavelength (0.9791 Å), was used to solve the structure (Supplementary Table 1). Nine Se atoms per asymmetric unit were located and refined using CNS²⁵. These sites were used to calculate phases in CNS with a mean figure of merit of 0.548 to 2.5 Å resolution. Following density modification by solvent flipping in CNS, RESOLVE^{26,27} was used to build residues 7–73 of one α -domain and residues 103–198 of one β -domain. The remaining residues were built manually into the density-modified experimental map using XFIT (D. E. McRee), and this initial model was refined as a rigid body in CNS²⁵. Iterative rounds of model building in XFIT and refinement in CNS followed at 2.5 Å resolution. The CNS protocol included simulated annealing against a maximum likelihood target and B-factor refinement; non-crystallographic symmetry restraints and sigma cutoff were not used. Topology files for Tris were obtained using HIC-Up. The final model was refined against a 1.8 Å resolution native Tris-Co(II)-HppE data set (Supplementary Table 1). This model contains residues 5–87, 102–158, 160–198 in molecule A and 6–87, 102–198 in molecule B (198 residues per molecule). There is no electron density for residues 88–101.

The refined Tris-Co(II)-HppE structure, including all modelled protein atoms, was used as a molecular replacement model to solve both the Fe(II)-HppE and the apo-HppE structure. The former was solved using AMORE²⁸ (correlation coefficient 0.58; R-factor 0.48) and the latter was solved using EPMR²⁹ (correlation coefficient 0.54; R-factor 0.46), in both cases using data from 15–4.0 Å resolution. Rigid body refinement was subsequently performed using CNS. The models were adjusted manually in XFIT and refined in CNS, using the same protocol as described above, to the highest available resolution (Supplementary Table 1). The final Fe(II)-HppE model contains residues 6–159, 161–198 in molecule A, 6–198 in molecule B, and 6–96, 100–198 in molecule C. The final apo-HppE model contains residues 6–95, 101–136, 140–157, 162–198 in molecule A and 6–97, 101–136, 141–158, 162–198 in molecule B (198 residues per molecule). The electron density was not interpretable for the remaining residues. The apo-HppE data were highly anisotropic, and therefore not of as good a quality as the other data sets that were used to refine HppE structures. As a result, the apo-HppE structure, with R-factors of 27.8% and 32.4% (R_{free}), is not as accurate as the other structures, but is useful in terms of a rough comparison with the metal-bound structures.

As S-HPP-Co(II)-HppE, form-1 S-HPP-Fe(II)-HppE, and Tris-Co(II)-HppE crystals are isomorphous (Supplementary Tables 1 and 2), the S-HPP-Co(II)-HppE, and S-HPP-Fe(II)-HppE structures were solved by rigid body refinement in CNS using only protein atoms from the refined Tris-Co(II)-HppE structure as an initial model. The 9 σ electron density peaks in the Co(II) structure and 6 σ peaks in the Fe(II) structure readily identified the location of the phosphonic moiety of the substrate, and allowed for the placement of the substrate in each model, respectively. Topology and parameter files for the substrate were obtained using HIC-Up. Manual adjustment of the model was done using XFIT and refinement was performed in CNS as described above. The final S-HPP-Co(II)-HppE model contains residues 6–90, 101–187, 190–198 in molecule A and 6–198 in molecule B (198 residues per molecule). The final form-1 S-HPP-Fe(II)-HppE model contains residues 7–87, 102–198 in molecule A and 7–87, 103–156, 160–198 in molecule B. The electron density for omitted residues was not sufficient for model building.

As form-2 S-HPP-Fe(II)-HppE and Fe(II)-HppE crystals are isomorphous (Supplementary Table 2), the form-2 S-HPP-Fe(II)-HppE structure was solved by rigid body refinement using protein atoms from the refined Fe(II)-HppE

structure as an initial model. The substrate was oriented on the basis the 10 σ electron density peaks that corresponded to substrate phosphorus atoms. Refinement was carried out as described above. The final form-2 S-HPP-Fe(II)-HppE model contains residues 7–198 in molecule A, and 6–198 in molecules B and C.

Simulated annealing composite omit maps were used to validate all HppE structures. For the six structures presented, all residues are in the allowed regions of the Ramachandran plot as calculated by PROCHECK³⁰: Fe(II)-HppE, 87.0% in the most favoured regions, 11.7% in additionally allowed regions and 1.3% in generously allowed regions; apo-HppE, 90.9% in the most favoured regions and 9.1% in additionally allowed regions; Tris-Co(II)-HppE, 90.2% in the most favoured regions, 9.5% in additionally allowed regions and 0.3% in generously allowed regions; S-HPP-Co(II)-HppE, 89.7% in the most favoured regions, 9.6% in additionally allowed regions and 0.6% in generously allowed regions; S-HPP-Fe(II)-HppE form-1, 90.7% in the most favoured regions, 8.9% in additionally allowed regions and 0.3% in generously allowed regions; S-HPP-Fe(II)-HppE form-2, 88.6% in the most favoured regions, 10.8% in additionally allowed regions and 0.6% in generously allowed regions; S-HPP-Co(II)-HppE, 89.7% in the most favoured regions, 9.6% in additionally allowed regions and 0.6% in generously allowed regions.

Received 28 December 2004; accepted 15 June 2005.

Published online 13 July 2005.

- Liu, P. *et al.* Protein purification and functional assignment of the epoxidase catalyzing the formation of fosfomycin. *J. Am. Chem. Soc.* **123**, 4619–4620 (2001).
- Lobel, B. Short term therapy for uncomplicated urinary tract infection today. Clinical outcome upholds the theories. *Int. J. Antimicrob. Agents* **22**, 85–87 (2003).
- Nakazawa, H., Kikuchi, Y., Honda, T., Isago, T. & Nozaki, M. Enhancement of antimicrobial effects of various antibiotics against methicillin-resistant *Staphylococcus aureus* (MRSA) by combination with fosfomycin. *J. Infect. Chemother.* **9**, 304–309 (2003).
- Cassone, M., Campanile, F., Pantosti, A., Venditti, M. & Stefani, S. Identification of a variant "Rome clone" of methicillin-resistant *Staphylococcus aureus* with decreased susceptibility to vancomycin, responsible for an outbreak in an intensive care unit. *Microb. Drug Resist.* **10**, 43–49 (2004).
- Hidaka, T. *et al.* Cloning and nucleotide sequence of fosfomycin biosynthetic genes of *Streptomyces wedmorensis*. *Mol. Gen. Genet.* **249**, 274–280 (1995).
- Hammerschmidt, F. Biosynthesis of natural-products with a P-C bond. 8. On the origin of the oxirane oxygen atom of fosfomycin in *Streptomyces fradiae*. *J. Chem. Soc. Perkin Trans. 1*, 18, 1993–1996 (1991).
- Liu, P. *et al.* Biochemical and spectroscopic studies on (S)-2-hydroxypropylphosphonic acid epoxidase: a novel mononuclear non-heme iron enzyme. *Biochemistry* **42**, 11577–11586 (2003).
- Guengerich, F. P. Cytochrome P450 oxidations in the generation of reactive electrophiles: epoxidation and related reactions. *Arch. Biochem. Biophys.* **409**, 59–71 (2003).
- Costas, M., Mehn, M. P., Jensen, M. P. & Que, L. Jr Dioxygen activation at mononuclear nonheme iron active sites: enzymes, models, and intermediates. *Chem. Rev.* **104**, 939–986 (2004).
- Dunwell, J. M. Cupins: a new superfamily of functionally diverse proteins that include germins and plant storage proteins. *Biotechnol. Genet. Eng. Rev.* **15**, 1–32 (1998).
- Zhao, Z. *et al.* Mechanistic studies of HPP epoxidase: configuration of the substrate governs its enzymatic fate. *Angew. Chem. Int. Edn Engl.* **41**, 4529–4532 (2002).
- Roach, P. L. *et al.* Structure of isopenicillin N synthase complexed with substrate and the mechanism of penicillin formation. *Nature* **387**, 827–830 (1997).
- Dunwell, J. M., Culham, A., Carter, C. E., Sosa-Aguirre, C. R. & Goodenough, P. W. Evolution of functional diversity in the cupin superfamily. *Trends Biochem. Sci.* **26**, 740–746 (2001).
- Dunwell, J. M., Purvis, A. & Khuri, S. Cupins: the most functionally diverse protein superfamily? *Phytochemistry* **65**, 7–17 (2004).
- Hausinger, R. P. Fe(II)/ α -ketoglutarate-dependent hydroxylases and related enzymes. *Crit. Rev. Biochem. Mol. Biol.* **39**, 21–68 (2004).
- Liu, P. *et al.* Oxygenase activity in the self-hydroxylation of (S)-2-hydroxypropylphosphonic acid epoxidase involved in fosfomycin biosynthesis. *J. Am. Chem. Soc.* **126**, 10306–10312 (2004).
- Ryle, M. J. *et al.* O₂- and α -ketoglutarate-dependent tyrosyl radical formation in TauD, an α -keto acid-dependent non-heme iron dioxygenase. *Biochemistry* **42**, 1854–1862 (2003).
- Roach, P. L., Schofield, C. J., Baldwin, J. E., Clifton, I. J. & Hajdu, J. Crystallization and preliminary X-ray diffraction studies on recombinant isopenicillin N synthase from *Aspergillus nidulans*. *Protein Sci.* **4**, 1007–1009 (1995).
- Valegård, K. *et al.* The structural basis of cephalosporin formation in a mononuclear ferrous enzyme. *Nature Struct. Mol. Biol.* **11**, 95–101 (2004).

20. Price, J. C., Barr, E. W., Tirupati, B., Bollinger, J. M. Jr & Krebs, C. The first direct characterization of a high-valent iron intermediate in the reaction of an α -ketoglutarate-dependent dioxygenase: a high-spin Fe^{IV} complex in taurine/ α -ketoglutarate dioxygenase (TauD) from *Escherichia coli*. *Biochemistry* **42**, 7497–7508 (2003).
21. Zhou, J. *et al.* Spectroscopic studies of substrate interactions with clavamate synthase 2, a multifunctional α -KG-dependent non-heme iron enzyme: correlation with mechanisms and reactivities. *J. Am. Chem. Soc.* **123**, 7388–7398 (2001).
22. Solomon, E. I. *et al.* Geometric and electronic structure/function correlations in non-heme iron enzymes. *Chem. Rev.* **100**, 235–350 (2000).
23. Roach, P. L. *et al.* Anaerobic crystallisation of an isopenicillin N synthase-Fe(II)-substrate complex demonstrated by X-ray studies. *Eur. J. Biochem.* **242**, 736–740 (1996).
24. Hegg, E. L. & Que, L. Jr The 2-His-1-carboxylate facial triad—an emerging structural motif in mononuclear non-heme iron(II) enzymes. *Eur. J. Biochem.* **250**, 625–629 (1997).
25. Brünger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
26. Terwilliger, T. C. Automated main-chain model building by template matching and iterative fragment extension. *Acta Crystallogr. D* **59**, 38–44 (2003).
27. Terwilliger, T. C. Maximum-likelihood density modification. *Acta Crystallogr. D* **56**, 965–972 (2000).
28. Navaza, J. AMoRe: an Automated Package for Molecular Replacement. *Acta Crystallogr.* **A50**, 164–182 (1994).
29. Kissinger, C. R., Gehlhaar, D. K. & Fogel, D. B. Rapid automated molecular replacement by evolutionary search. *Acta Crystallogr. D* **55**, 484–491 (1999).
30. Laskowski, R., MacArthur, M., Moss, D. & Thornton, J. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291 (1993).
31. DeLano, W. L. *The PyMOL Molecular Graphics System* (DeLano Scientific, San Carlos, California, 2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This research is supported in part by the National Institutes of Health (C.L.D. and H-W.L.), the National Institute of Environmental Health Sciences (L.J.H.), the Searle Scholars Program (C.L.D.), Alfred P. Sloan Foundation (C.L.D.), and a Lester Wolfe Predoctoral Fellowship (L.J.H.). Synchrotron facilities are funded by the NIH National Center of Research Resources (Advanced Photon Source NE-CAT 8BM) and the US Department of Energy, Division of Materials Sciences and Division of Chemical Sciences (National Synchrotron Light Source, Brookhaven National Laboratory).

Author Information Atomic coordinates for the following structures have been deposited in the Protein Data Bank: apo-HppE (1ZZ6), Fe(II)-HppE (1ZZ9), form-1 S-HPP-Fe(II)-HppE (1ZZ7), form-2 S-HPP-Fe(II)-HppE (1ZZ8), Tris-Co(II)-HppE (1ZZC) and S-HPP-Co(II)-HppE (1ZZB). Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.L.D. (cdrennan@mit.edu).

The afterglow of GRB 050709 and the nature of the short-hard γ -ray bursts

D. B. Fox^{1,2}, D. A. Frail³, P. A. Price⁴, S. R. Kulkarni¹, E. Berger⁵, T. Piran^{1,6}, A. M. Soderberg¹, S. B. Cenko¹, P. B. Cameron¹, A. Gal-Yam¹, M. M. Kasliwal¹, D.-S. Moon¹, F. A. Harrison¹, E. Nakar¹, B. P. Schmidt⁷, B. Penprase⁸, R. A. Chevalier⁹, P. Kumar¹⁰, K. Roth¹¹, D. Watson¹², B. L. Lee¹³, S. Shetman⁵, M. M. Phillips⁵, M. Roth⁵, P. J. McCarthy⁵, M. Rauch⁵, L. Cowie⁴, B. A. Peterson⁷, J. Rich⁷, N. Kawai¹⁴, K. Aoki¹⁵, G. Kosugi¹⁵, T. Totani¹⁶, H.-S. Park¹⁷, A. MacFadyen¹⁸ & K. C. Hurley¹⁹

The final chapter in the long-standing mystery of the γ -ray bursts (GRBs) centres on the origin of the short-hard class of bursts, which are suspected on theoretical grounds to result from the coalescence of neutron-star or black-hole binary systems. Numerous searches for the afterglows of short-hard bursts have been made, galvanized by the revolution in our understanding of long-duration GRBs that followed the discovery in 1997 of their broadband (X-ray, optical and radio) afterglow emission. Here we present the discovery of the X-ray afterglow of a short-hard burst, GRB 050709, whose accurate position allows us to associate it unambiguously with a star-forming galaxy at redshift $z = 0.160$, and whose optical lightcurve definitively excludes a supernova association. Together with results from three other recent short-hard bursts, this suggests that short-hard bursts release much less energy than the long-duration GRBs. Models requiring young stellar populations, such as magnetars and collapsars, are ruled out, while coalescing degenerate binaries remain the most promising progenitor candidates.

High-energy transients remain at the frontier of astrophysics research because they probe extreme physical regimes of matter, gravity and energy density. Soft γ -ray repeaters (SGRs) combine matter at supra-nuclear densities with magnetic fields in excess of 10^{15} G, while long-duration GRBs, which probably herald the birth of stellar-mass black holes, drive ultra-relativistic outflows and power the brightest explosions in the Universe. Progress in understanding new classes of high-energy transient events has typically required a multi-wavelength approach; in particular, the identification of longer-wavelength counterparts enables precise localization and detailed studies. It was the discovery of the slow-fading ‘afterglow’ emission of long-duration GRBs that enabled their subarcsecond localization, the measurement of their redshifts, the identification of their star-forming host galaxies, the quantification of their energy scale, and ultimately, established their connection to the deaths of massive stars (see ref. 1 for a review).

The nature of the short-hard γ -ray bursts (SHBs) has been an outstanding mystery of high-energy astrophysics for more than 30 years. The SHBs comprise about 30% of the GRB population at the Burst and Transient Source Experiment (BATSE) threshold and have typical durations of 0.3 s and peak energies of the order of 350 keV, with power-law tails extending to higher energies². Despite extensive searches³, no short-hard burst has yet been sufficiently well-localized to ascertain its origins. Historically, γ -ray satellites were either not sensitive to SHBs or provided positions that were too crude or too

delayed to enable deep searches. A significant breakthrough came when the X-ray telescope (XRT) on the recently launched Swift satellite detected the rapidly fading afterglow of GRB 050509B and localized it to a circular region of radius 9.3 arcsec. Within this X-ray localization there are nearly 50 objects identified in Hubble Space Telescope (HST) images⁴, the brightest of which by far is an elliptical galaxy at $z = 0.2248$ that has been proposed as the likely host galaxy of this burst^{5,6}.

Even without any SHB distance determinations, the isotropic sky distribution and non-euclidean brightness distribution of the SHBs suggest a cosmological origin^{2,7}, fuelling speculation that SHBs are the result of the coalescence of compact object (neutron star–neutron star or neutron star–black hole) binaries⁸. Theoretical estimates yield merger rates⁹ that can easily accommodate the observed burst rate, with engine lifetimes and energy releases roughly consistent with the burst properties for a cosmological population. Nonetheless, without any detailed knowledge of their distances, energetics and environments, younger progenitor populations such as magnetars and collapsars cannot be ruled out. If the coalescence model is correct, the SHBs will be a primary source population for Laser Interferometer Gravitational Wave Observatory (LIGO) and other ground-based gravitational wave detectors. The SHBs thus promise to provide a crucial test-bed for theories of strong-field gravity, the nuclear equation of state and the formation of black holes.

¹Division of Physics, Mathematics and Astronomy, California Institute of Technology, Pasadena, California 91125, USA. ²Department of Astronomy and Astrophysics, 525 Davey Laboratory, Pennsylvania State University, University Park, Pennsylvania 16802, USA. ³National Radio Astronomy Observatory, PO Box O, Socorro, New Mexico 87801, USA.

⁴University of Hawaii, Institute for Astronomy, 2680 Woodlawn Drive, Honolulu, Hawaii 96822, USA. ⁵Carnegie Observatories, 813 Santa Barbara Street, Pasadena, California 91101, USA. ⁶Racah Institute for Physics, The Hebrew University, Jerusalem 91904, Israel. ⁷Research School of Astronomy and Astrophysics, The Australian National University, Weston Creek, ACT 2611, Australia. ⁸Pomona College, 610 North College Avenue, Claremont, California 91711, USA. ⁹Department of Astronomy, University of Virginia, PO Box 3818, Charlottesville, Virginia 22903, USA. ¹⁰Astronomy Department, University of Texas, Austin, Texas 78731, USA. ¹¹Gemini Observatory, 670 North A'ohoku Place, Hilo, Hawaii 96720, USA. ¹²Niels Bohr Institute, University of Copenhagen, Juliane Maries Vej 30, DK-2100 Copenhagen, Denmark. ¹³Department of Astronomy and Astrophysics, University of Toronto, Toronto, Ontario M5S 3H8, Canada. ¹⁴Department of Physics, Tokyo Institute of Technology, Ookayama 2-12-1, Meguro-ku, Tokyo 152-8551, Japan. ¹⁵Subaru Telescope, National Astronomical Observatory of Japan, 650 N. A'ohoku Place, Hilo, Hawaii 96720, USA. ¹⁶Department of Astronomy, School of Science, Kyoto University, Sakyo-ku, Kyoto 606-8502, Japan. ¹⁷Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, California 94550, USA. ¹⁸Institute for Advanced Study, Princeton, New Jersey 08540, USA. ¹⁹Space Sciences Laboratory, University of California, Berkeley, California 94720, USA.

Discovery of the X-ray afterglow

Upon receiving notification of the localization¹⁰ of the short-hard burst GRB 050709 by the High-Energy Transient Explorer (HETE), we initiated observations with the Chandra X-ray Observatory as part of our approved programme for the SHBs, observing the 81-arcsec error circle with the Advanced charge-coupled device (CCD) Imaging Spectrometer¹¹. A total of 38.4 ks of good data were obtained after excluding intervals of background flaring activity, at a mean epoch of 2.52 days after the burst (Table 1). We detected two sources in the HETE error circle: a faint and resolved (or double) source and a bright point source (see Fig. 1). The faint source is well detected at low energies (0.3–2.0 keV band), has a flux of $3.0 \times 10^{-15} \text{ erg cm}^{-2} \text{ s}^{-1}$ (1–5 keV band) and coincides with a catalogued radio source from the National Radio Astronomy Observatory/Very Large Array (VLA) 20-cm Sky Survey¹² (NVSS).

The bright point source has 49.5 ± 8.8 counts in the 0.3–8 keV band, corresponding to an X-ray flux of $3.5 \times 10^{-15} \text{ erg cm}^{-2} \text{ s}^{-1}$ (1–5 keV band) for the best-fit power-law spectrum (photon index $\alpha = 2.24 \pm 0.35$, with column density fixed to the expected Galactic hydrogen column density, $N_{\text{H}} = 1.2 \times 10^{20} \text{ cm}^{-2}$). The source is located at $\alpha = 23:01:26.96$, $\delta = -38:58:39.5$ (J2000). This position has been corrected by 0.4 arcsec from the native Chandra astrometry using three X-ray sources coincident with stars in the United States Naval Observatory (USNO) B1.0 catalogue. We estimate the 90% confidence radius is 0.5 arcsecond.

We proposed¹³ the brighter source as the X-ray afterglow of GRB 050709. We also noted that it was offset by about one arcsecond from a faint ($R \approx 20.5$ mag) source visible in the Digitized Sky Survey,

plausibly its host galaxy. Making use of the Chandra position, we found a marginal detection of the X-ray source in earlier observations made by the Swift XRT (Table 1).

We then executed an 18-ks follow-up observation (mean epoch 16.0 days post-burst) with an identical observatory configuration. These data showed that, at a 99.7% confidence level, the X-ray source had faded by more than a factor of two. Inspecting the eleven events within a 1.5-arcsec radius of the X-ray afterglow position, we found that nine events occurred within the first third of the observation. This is in contrast to the first epoch, which exhibits a roughly uniform count rate over the observation. A Kolmogorov–Smirnov test demonstrates that the second epoch arrival times are inconsistent with a steady event rate at 99.9% confidence. We therefore suggest that during the first ~ 6 ks of this observation the source was in a ‘flaring’ state, roughly an order of magnitude brighter than during the remainder of the observation.

In Table 1 we give the mean epochs and X-ray fluxes for the flaring and quiescent portions of this observation. Using the quiescent flux, which represents only a marginal detection of emission (90% confidence), we find a temporal decay index in the X-ray band of $\alpha_{\text{X}} \approx -1$ for the interval from 2.5 to 16 days after the burst.

The discovery of this flaring behaviour suggests that even 16 days after the burst, the afterglow is still subject to new energy inputs. The sudden cessation of the flare represents a small fraction of the time since the burst, indicating that the flaring must arise from a source physically distinct from the fading afterglow. We suggest that the flare arises from ongoing activity of the central engine, in analogy to the bright X-ray flares observed from several long-duration Swift GRBs¹⁴.

Table 1 | Observations of the afterglow of GRB 050709

Facility	UT date in 2005	UT	ΔT (days)	Band	Exp	Flux
Swope-40*	11 July	04:20	1.24	i'	3×600 s	>20.5 mag
Swope-40*	11 July	10:19	1.49	i'	3×600 s	...
Du Pont-100*	13 July	08:33	3.41	R	3×600 s	(21.05 ± 0.15) mag
Subaru	15 July	14:06	5.6	K'	270×20 s	22.1 ± 0.7 mag
Subaru	26 July	13:19	16.6	K'	360×20 s	(19.2 ± 0.1)
VLA†	11 July	12:14	1.6	8.46 GHz	6,660 s	$<76.6 \mu\text{Jy}$
VLA†	12 July	09:36	2.5	8.46 GHz	1,265 s	$<75.8 \mu\text{Jy}$
VLA†	14 July	11:31	4.5	8.46 GHz	6,055 s	$<49.4 \mu\text{Jy}$
VLA†	17 July	10:48	7.5	8.46 GHz	6,490 s	$<26.6 \mu\text{Jy}$
HETE WXM	09 July	22:38	100 s	5 keV	10 s	0.80 ± 0.14 mJy
Chandra‡	12 July	11:15	2.5	5 keV	38.4 ks	0.15 ± 0.02 nJy
Chandra‡	25 July	21:47	16.0	5 keV	6.1 ks	0.18 ± 0.06 nJy
Chandra‡	26 July	00:18	16.1	5 keV	12.1 ks	$0.015^{+0.027}_{-0.007}$ nJy
Swift§	11 July	13:07	1.6	5 keV	15.2 ks	0.26 ± 0.14 nJy
Swift§	12 July	10:13	2.4	5 keV	5.0 ks	<0.2 nJy
Swift§	13 July	02:32	3.2	5 keV	6.4 ks	<0.1 nJy
Swift§	14 July	05:26	4.3	5 keV	2.1 ks	<0.8 nJy
HST	15 July	13:49	5.6	F814W	6,360 s	25.08 ± 0.02 mag
HST	19 July	17:11	9.8	F814W	6,360 s	25.84 ± 0.05 mag
HST	28 July	13:48	18.6	F814W	6,360 s	27.81 ± 0.27 mag
HST	13 August	15:17	34.7	F814W	6,360 s	>28.1 mag

We carried out all observations summarized in this table except those of HETE (from ref. 10) and Swift (reported in ref. 38). The HETE point represents the average flux in the ‘soft flare’ that occurs 100 s after the main burst. We have made an independent reduction of the Swift XRT data. WXM, Wide-field X-ray Monitor.

* Optical and near-infrared. The optical and near-infrared fluxes do not contain a correction for Galactic extinction, which is expected to be rather small in this direction³⁹: $E(B-V) = 0.012$ mag. Near-infrared observations were made with the Cool Infrared Spectrograph and Camera for OH Suppression⁴⁰. No flux is reported for the Swope-40 second epoch because the image was subtracted from the first epoch in order to search for the afterglow. Host fluxes in the table are given in parentheses.

† Radio. The VLA data were taken in standard continuum mode with a bandwidth of 2×50 MHz centred in the 8.46 GHz band. We used 3C48 for flux calibration and phase referencing was performed against calibrator J2257 – 364. Data were reduced using standard packages within the Astronomical Image Processing System. Within the HETE error circle we find a single source with constant flux ($644 \pm 24 \mu\text{Jy}$) at coordinates $\alpha = 23:01:32.1 \pm 0.003$, $\delta = -38:59:26.8 \pm 0.1$ (J2000). This source is coincident with the catalogued extended NVSS source and resolved Chandra source. Upper limits are quoted at 2σ or 95.5% confidence.

‡ Chandra X-ray Observatory. Our basic data reduction procedures are described in the text. We use a custom pipeline composed of CIAO 3.2.1 tools (<http://cxc.harvard.edu/ciao/>) to run a full ‘wavdetect’ analysis on images constructed from the 0.3–2.0 keV, 2.0–8.0 keV and 0.3–8.0 keV bands separately, and then merge the resulting source catalogues. For spectral fits, photons are extracted from a $1.5''$ -radius aperture, and fits are made using XSPEC 12.0 (<http://heasarc.gsfc.nasa.gov/docs/xanadu/xspec/>). The second observation has been divided into ‘flaring’ and ‘quiescent’ intervals; see text for details.

§ Swift X-ray Telescope. The photometry was done using a circular region of radius 0.8 arcmin centred on the X-ray afterglow position. The background was estimated from the entire XRT image. The detection confidence level from Poisson statistics is 2.3σ for the first observation; the remaining observations provide only upper limits (quoted at 2σ or 95.5% confidence). Photon count limits are converted to flux densities using our spectral fit to the first-epoch Chandra data; see text for details.

|| Hubble Space Telescope. Data were obtained with the ACS instrument aboard HST¹⁷. Each epoch consisted of a series of exposures in the F814W (I-band) filter with a total integration time of 6,360 s. The images were processed using Archive ‘on-the-fly’ processing, drizzled⁴¹ to the native pixel scale of 0.05 arcsecond, and combined. We subtracted the fourth-epoch image from each previous epoch and performed photometry using a $0.15''$ (3-pixel) aperture. The magnitudes, quoted in the ‘AB’ system, are corrected for finite aperture (0.32 mag) and imperfect charge transfer efficiency (0.01 mag). Errors are estimated by photometering multiple background regions in the subtracted images using the same aperture. The limit on afterglow flux in the fourth epoch is derived by subtracting point sources of decreasing flux from the image until a flux deficit at the afterglow position is no longer readily discernible.

Optical afterglow and host galaxy

In addition to our Chandra observations, we conducted an extensive ground-based campaign on GRB 050709 at radio, optical and near-infrared wavelengths using the VLA, the 40-inch Swope and 100-inch Du Pont telescopes at Las Campanas Observatory, and the 8.2-m Subaru Telescope on Mauna Kea. A complete list of these observations is given in Table 1, along with upper limits on the flux of the afterglow at these epochs and measurements of the host galaxy brightness.

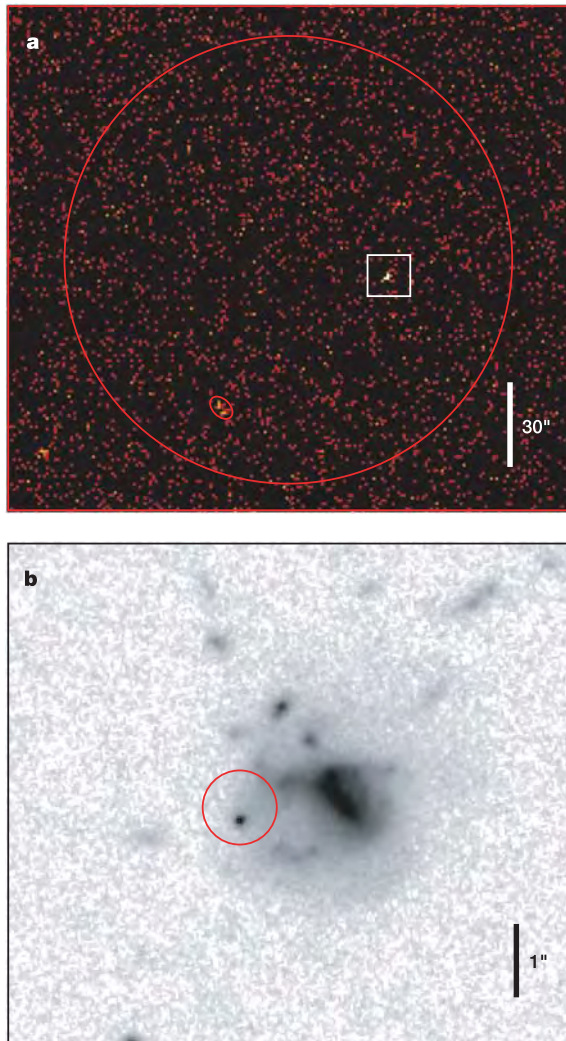


Figure 1 | HST and Chandra X-ray Observatory images of the afterglow and environs of GRB 050709. **a**, The Chandra (0.3–8.0 keV) image of the field from our observation of 2005 12.5 July UT. The large red circle is the HETE localization region, 81 arcsec in radius; north is up, east is to the left, and the scale of the image is indicated. A red ellipse indicates the faint X-ray source that we identify with an NVSS catalogue object; the bright point source in the boxed region is the afterglow of GRB 050709. **b**, Close-up of the region surrounding the X-ray afterglow, in a co-addition of all our HST data; the red circle is the Chandra localization region, 0.5 arcsec in radius. A point source is visible within this region; the source is observed to fade over the course of our HST observations, and we identify it as the optical afterglow of GRB 050709. The irregular galaxy to its west is the proposed $z = 0.16$ host galaxy. We use the GALFIT software⁴³ to fit the radial surface brightness distribution of the host galaxy, with the Sérsic concentration parameter, n , and the effective radius, r_e , as free parameters. We find a best-fit solution ($\chi^2 \approx 3$ per degree of freedom) with $n = 1.1$ and $r_e = 0.75''$. At $z = 0.16$ the afterglow's 1.38-arcsec offset from the brightest region of this galaxy corresponds to 3.8 kpc in projection, or $1.8r_e$.

Our Chandra afterglow candidate was found to be coincident with a point-like optical source¹⁵, distinct from the candidate host galaxy which faded in a manner consistent with the optical afterglows of long-duration GRBs¹⁶. We undertook spectroscopy of the candidate host galaxy with the Gemini Multi-Object Spectrograph (GMOS) on the Gemini North telescope, and find it to be a star-forming galaxy at redshift $z = 0.160$ (Fig. 2).

We also triggered a sequence of HST observations with the Advanced Camera for Surveys¹⁷ (ACS). Within the Chandra error circle we find a single bright, fading, point-like source, the unambiguous optical afterglow of GRB 050709; our HST photometry is presented in Table 1. Expressing the afterglow decay as a power-law (flux $\propto t^\alpha$) between each epoch, we find power-law indices of $\alpha_{12} = -1.25 \pm 0.09$ between epochs 1 and 2, $\alpha_{23} = -2.83 \pm 0.39$ between epochs 2 and 3, and $\alpha_{34} < -0.43$ between epochs 3 and 4. Our observation of a break in the decay is definitive; the HST data are inconsistent with a single power-law decay at the 3.7σ level.

As can be seen from Fig. 1, the optical afterglow of GRB 050709 is superposed on the outskirts of the $z = 0.16$ candidate host galaxy. This precise localization, the first subarcsecond localization for any SHB, unambiguously associates GRB 050709 with the $z = 0.16$ galaxy. Thus we show here, to our knowledge for the first time, that some SHBs arise in low-redshift star-forming galaxies.

The morphology of the host galaxy is irregular, typical of star-forming galaxies. We fitted the radial light profile and found that it is well described by an exponential disk with scale length $r_e = 0.75''$. The afterglow is situated 1.38 arcsec or $1.8r_e$ from the brightest central region of the galaxy, corresponding to a physical offset of 3.8 kpc. From the detected emission lines we derive a star-formation rate of $0.2M_\odot \text{ yr}^{-1}$ (a lower limit after allowing for extinction). By comparison, long-duration GRBs are found exclusively in late-type (star-forming) host galaxies¹⁸, and with a somewhat smaller median offset¹⁸ of $1.0r_e$.

Burst and afterglow energetics

At a redshift of $z = 0.16$, the isotropic-equivalent energy release in γ -rays¹⁰ over the 25–2,000 keV band is $E_{\gamma, \text{iso}} = 6.9^{+1.0}_{-0.5} \times 10^{49}$ erg and

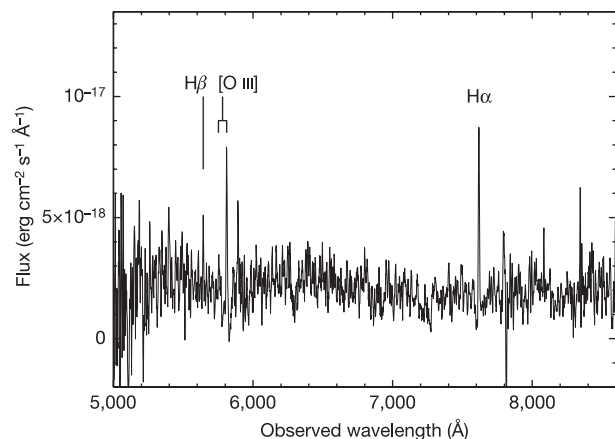


Figure 2 | Spectrum of the host galaxy of GRB 050709. These observations were taken with the GMOS on the Gemini North telescope under poor seeing and non-photometric conditions. We obtained a total integration of 3,303 s in four exposures before closing owing to the onset of twilight. The spectra were processed with the CMOS data reduction package in IRAF, sky-subtracted, combined and extracted, and smoothed with a 7 Å boxcar. Flux calibration was performed using an observation of Hiltner 600 observed with a similar instrument set-up in a previous lunation; hence, the flux scale shown is indicative rather than absolute. Three emission lines are readily apparent ($H\alpha$, $[O \text{ III}]$ at $\lambda = 5,007 \text{ Å}$ and $H\beta$) yielding a mean redshift of $z = 0.160 \pm 0.001$. The observed $H\alpha/H\beta$ flux ratio of 3.1 indicates that the global galaxy extinction is small ($A_V = 0.1 \text{ mag}$).

the peak luminosity is $L_p = (1.1 \pm 0.5) \times 10^{51} \text{ erg s}^{-1}$ (here and throughout this paper, we adopt a flat cosmology with $H_0 = 71 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_M = 0.27$ and $\Omega_\Lambda = 0.73$). The burst of γ -rays is followed by an X-ray flare, detected by HETE from 25 s to 130 s after the burst¹⁰. The fluence of this X-ray flare is twice as much as that of the γ -ray burst itself. Thus, the total isotropic energy release in the first few hundred seconds is $E_{\text{iso}} \approx 10^{50} \text{ erg}$, two orders of magnitude smaller than that seen in long-duration bursts¹⁹.

Figure 3 presents our observations of the GRB 050709 afterglow. The steepening power-law decay seen in our HST observations, a familiar feature of long-duration GRB light curves, is usually explained as arising from the collimated or jet-like nature of the ejecta²⁰. This is to our knowledge the first observation of such a light-curve 'jet break' for an SHB, although the steep ($\alpha \approx -2$) decay of the GRB 050724 afterglow is suggestive²¹. The epoch of steepening, $t_b \approx 10$ days, can be related to the opening angle of the jet, θ_j , as follows¹⁹ (in radians):

$$\theta_j = 0.076 \left(\frac{t_b}{1 \text{ day}} \right)^{3/8} \left(\frac{1+z}{2} \right)^{-3/8} \times \left(\frac{\eta_\gamma}{0.2} \right)^{1/8} \left(\frac{n}{1 \text{ cm}^{-3}} \right)^{1/8} \left[\frac{E_{\gamma, \text{iso}}}{10^{53} \text{ erg}} \right]^{-1/8} \quad (1)$$

Here n is the circumburst particle density and η_γ is the ratio of the

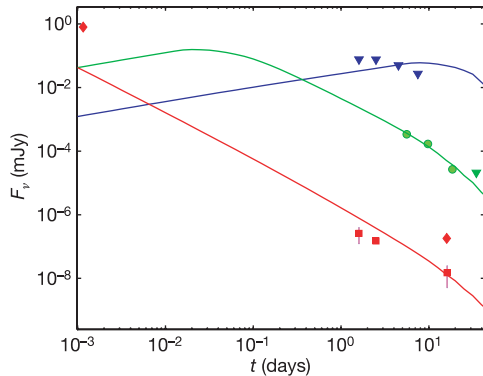


Figure 3 | Observations of the GRB 050709 afterglow and illustrative models. The X-ray (red), optical (green) and radio (blue) data taken from Table 1. These multi-wavelength observations can be marginally accommodated within a standard external-shock afterglow model. The best-fit parameters are $E = 5 \times 10^{48} \text{ erg}$, $\epsilon_e = \epsilon_B = 1/3$, $n = 0.01$, $p = 2.5$ and $\theta_j = 0.35$. However, even this best-fit predicts a radio flux which is slightly higher (a factor of two) than our upper limits and it requires an uncommonly high value for ϵ_B . The latter is a concern because we do not expect the microphysics of the afterglow to differ between the short- and long-duration GRB populations, although one cannot exclude this possibility. The early HETE X-ray observation and the late Chandra X-ray flare cannot be explained by the model: both are too bright. The early point could be a flare resulting from an energy injection to the external shock or a long-lasting activity of the source, as seen recently by the Swift/XRT in a long burst¹⁴. The late flare, however, is a phenomenon unusual in short bursts and not observed so far in any long burst. The duration of the flare (~ 0.1 day) corresponds to an emitting region at a size of 10^{14} cm 16 days after the burst. At this time the external shock is mildly relativistic and its width is $\sim 10^{16} \text{ cm}$. The (isotropic equivalent) energy emitted in this flare is $\sim 10^{45} \text{ erg}$ in the X-ray spectrum alone. These different length scales, together with the brightness of the flare, exclude the possibility that the flare is produced in a simple external shock and indicates a process that has not been observed so far in long bursts. Most probably, this process involves an activity of the source at $t \approx 16$ days that is 10^7 times the duration of the burst. Combined with the unique X-ray light curve of the short-hard burst GRB 050724 (showing an early, very steep decay and a later brightening) these observations suggest that we should keep an open mind for the possibility that SHB afterglows are a very different phenomenon from the afterglows of long GRBs.

radiated energy to the energy in relativistic ejecta. Setting $E_{\gamma, \text{iso}} = 10^{50} \text{ erg}$ and $n = 10^{-2} \text{ cm}^{-3}$ we obtain $\theta_j = 0.25$ and a beaming fraction, $f_b = 1 - \cos(\theta) \approx 0.03$. With this value of f_b , the energy released in the first few hundred seconds is $3 \times 10^{48} \text{ erg}$ —two orders of magnitude less than that inferred for long-duration GRBs¹⁹.

We interpret the fading optical and X-ray emission to be the afterglow of GRB 050709. The afterglow phenomenon has been studied both observationally and theoretically in the context of long-duration GRBs¹. The emission arises from ambient material shocked by the relativistic ejecta. The broadband spectral index between the optical and X-ray bands at late times (Chandra and HST observations) is $\beta_{\text{OX}} = -1.1$; because the temporal power-law index is $\alpha_o = -1.25$ in the optical prior to the jet break, the electron power-law index is derived to be $p = 2.7$, and we find that the synchrotron cooling frequency is between the optical and X-ray bands at this time.

The isotropic-equivalent X-ray luminosity $L_{X, \text{iso}}$ of the afterglow at a fixed time after the burst serves as a useful proxy²² for the kinetic energy of the ejecta in the afterglow phase. Extrapolating back from the first Chandra flux to a fiducial time of 10 h post-burst, assuming a $t^{-1.3}$ decay (Fig. 3), we find $L_{X, \text{iso}} \approx 2 \times 10^{42} \text{ erg s}^{-1}$, at least three orders of magnitude smaller than is typical for the afterglows of long-duration GRBs²³.

Limits on an associated supernova

At $z = 0.16$, the distance modulus of GRB 050709 is $(m - M) = 39.4$ mag. Our HST data (Fig. 1) are consistent with a pure afterglow evolution, and show no evidence for a supernova. We detected the afterglow in optical light, so we are in the unique position (for SHBs) of being able to exclude any role for extinction in suppressing the light of an associated supernova. We estimate our sensitivity limit for a supernova bump as the faintest magnitude we observe, $m_{\text{SN}} > 27.5$ mag. Converting to absolute magnitudes and applying a k -correction (from F814W AB to Vega magnitudes) of -0.12 mag, we estimate $M_R > -12.0$ mag for any associated supernova at an age of 16 days. This limit is fainter than any supernova yet detected in the nearby Universe.

Numerical studies of SHBs²⁴ predict a modest associated nova-type event, much fainter than the average supernova. Quantitative

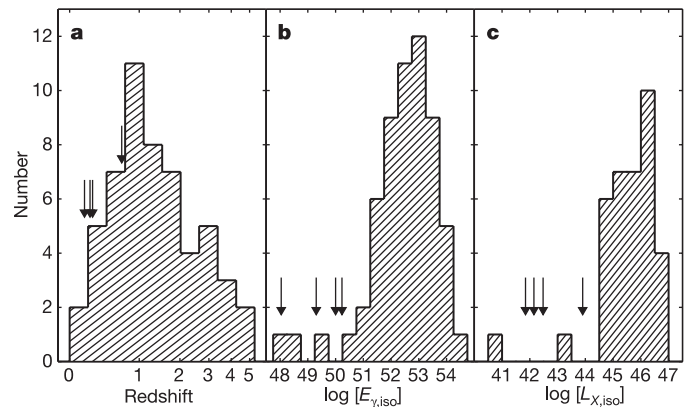


Figure 4 | Physical properties of the afterglows of long-duration GRBs (histograms) and SHBs (arrows). a–c, For GRBs with known redshifts, their distribution in redshift (a), isotropic-equivalent gamma-ray luminosity ($\log[E_{\gamma, \text{iso}}]$, b), and isotropic-equivalent X-ray luminosity at 10 h after the burst ($\log[L_{X, \text{iso}}]$, c) are shown. Arrows indicate the values of these properties for the four SHBs with afterglow detections, GRB 050509B, GRB 050709, GRB 050724 and GRB 050813 (ordered from lowest to highest, in each panel). The contrast between the physical properties of the two burst populations is noticeable. GRB properties are from refs 23 and 44–46. See Table 2 for SHB references.

predictions for the ejecta mass, speed and composition span an extremely broad range, however, so that the situation is ripe for observational inputs.

In the absence of heat input after the explosion, adiabatic expansion of any sub-relativistic ejecta results in very rapid fading (ref. 25, and S.R.K., manuscript in preparation). Continued heat input might be provided by the decay of radioactive nuclei (including nickel)²⁵, decay of free neutrons, or a long-lived central engine (S.R.K., manuscript in preparation). Any such heat input will be reprocessed to lower energies—mainly via electron Thompson scattering—provided the ejecta are dense enough. The high sensitivity of current facilities makes optical wavelengths the band of choice for these searches.

Our observations constrain the kinetic energy of slow ejecta in these models (S.R.K., manuscript in preparation) to be less than 10^{49} erg, provided that the ejecta velocity $v \leq 0.02c$. The current limit arises entirely from the first-epoch HST data. Sensitive optical data taken at earlier times would have provided stronger constraints and for higher ejecta speeds.

An alternative source of heat could be luminosity from a long-lived central engine. The HST observations constrain this luminosity to be $L \leq 10^{41}$ erg s⁻¹. We note that the X-ray flare in the second-epoch Chandra data has a similar luminosity.

Properties of SHBs

The offsets of GRB 050709 and GBB 050724 from their proposed host galaxies are small enough that their associations can be considered secure. These two associations, in turn, strengthen the case for identification of GRB 050509B (localized to a 9.3-arcsec radius⁶) with the redshift $z = 0.225$ galaxy⁵ and galaxy cluster⁶ that have been proposed to host this burst. We can thus set the physical scale for the energetics of all four bursts. Table 2 presents the properties of these SHBs, along with some properties of their host galaxies. Figure 4 places these bursts—the only known SHB afterglows—in the context of the set of long-duration GRBs with known redshifts.

In Table 2 we display the peak luminosities of the four SHBs, extrapolated to the full BATSE band for comparison with results from that experiment. All four values are approximately $L_{\gamma, \text{peak}} \approx 10^{50}$ erg s⁻¹. The similarity of the redshifts and peak luminosities of the SHBs suggests that they arise from a single source population.

Next, from Fig. 4 we see, relative to the long-duration GRBs, that SHBs are located at lower redshift, emit less energy in γ -rays, and possess a less-energetic afterglow. This behaviour is broadly consistent with conclusions from earlier statistical studies^{7,9,26,27}.

A closer distance scale for the SHBs is consistent with the value of $\langle V/V_{\text{max}} \rangle \approx 0.4$ for BATSE SHBs^{9,28}, which is significantly higher than the BATSE value for long-duration GRBs. In particular, the BATSE SHB $\langle V/V_{\text{max}} \rangle$ value is consistent with a spatially uniform distribution of standard candles out to $z_{\text{max}} \approx 0.4$ (ref. 29) and with peak luminosities of $L_{\gamma, \text{peak}} \approx 10^{50}$ erg s⁻¹, as observed for these four bursts.

In Table 2 we also summarize selected properties of the four SHB host galaxies. The proposed host of GRB 050509B is a large elliptical galaxy at redshift $z = 0.2248$ (ref. 5), with luminosity $L \approx 1.5L_*$ (ref. 30) where L_* is the characteristic luminosity of galaxies in the field as parameterized by the Schechter luminosity function. The galaxy has a very low ongoing star-formation rate of $< 0.1 M_{\odot} \text{ yr}^{-1}$, where $M_{\odot}$ is the mass of the Sun. The elliptical host galaxy of GRB 050724 (which has also been localized to sub-arcsecond precision) is a luminous ($L = 1.6L_*$) elliptical galaxy with a star-formation rate $< 0.02 M_{\odot} \text{ yr}^{-1}$ (ref. 21).

In contrast, the morphology and spectrum of the host galaxy of GRB 050709 (Fig. 1) indicates that the host is an irregular, late-type galaxy with a significant star-formation rate, $\sim 0.2 M_{\odot} \text{ yr}^{-1}$, and a luminosity much smaller than the other SHB hosts, $L \approx 0.10L_*$; so the galaxy is forming roughly as many stars per unit stellar mass as the Milky Way.

The association of SHBs with both star-forming galaxies and early-type ellipticals is reminiscent of the diversity of host galaxies of type Ia supernovae. As with the type Ia supernova, this dichotomy may indicate a class of progenitors with an extremely wide range of lifetimes between formation and explosion, with some systems living for many gigayears. However, even though type Ia supernovae occur in elliptical galaxies, the rate of such events is higher in late-type, star-forming galaxies, and the majority of type Ia events in the local Universe occur in late-type blue galaxies³¹. The trend emerging for the SHBs, in which the ‘majority’ of events (here, perhaps three of four events) occur in elliptical galaxies, could indicate that the progenitor systems are even longer-lived than those of type Ia supernovae.

Nature of the SHBs

Our observations of GRB 050709 support the view—until recently founded only on the basis of their prompt γ -ray emissions—that the SHBs are a different population from the long-duration GRBs (Fig. 4). They are lower-energy explosions, with a correspondingly less-energetic relativistic blast wave, and they are found at significantly closer distances. At the same time, the similarity of the SHB

Table 2 | Physical properties of short-hard bursts and their host galaxies

Property	GRB 050509B	GRB 050709	GRB 050724	GRB 050813
Redshift	0.225	0.160	0.258	0.722
T_{90}	40 ± 4 ms	70 ± 10 ms	3 ± 1 s	0.6 ± 0.1 s
Fluence (erg cm ⁻²)	9.5×10^{-9}	2.9×10^{-7}	6.3×10^{-7}	1.2×10^{-7}
Fluence band	15–350 keV	30–400 keV	15–350 keV	15–350 keV
$E_{\gamma, \text{iso}}$ (erg)	4.5×10^{48}	6.9×10^{49}	4.0×10^{50}	6.5×10^{50}
$L_{\gamma, \text{peak}}$ (erg s ⁻¹)	1.4×10^{50}	1.1×10^{51}	1.7×10^{50}	1.9×10^{51}
L_X (erg s ⁻¹)	$< 7 \times 10^{41}$	3×10^{42}	8×10^{43}	9×10^{43}
f_b	—	0.03	0.01	—
E_{γ} (erg)	$< 4.5 \times 10^{48}$	2.1×10^{48}	4.0×10^{48}	$< 6.5 \times 10^{50}$
Host L/L_*	1.5	0.10	1.6	—
Host star-formation rate ($M_{\odot} \text{ yr}^{-1}$)	< 0.1	0.2	< 0.03	—
Offset (kpc)	44^{+12}_{-23}	3.8	2.6	—
Offset (r/r_e)	13^{+3}_{-7}	1.8	0.4	—
Supernova limit, M_R (mag)	> -13.0	> -12.0	—	—
References	4, 5, 6	10; this work	21, 42	*

In order, the table shows: the burst redshift; the duration of the 90%-inclusive interval of high-energy emission (T_{90}); the measured burst fluence and corresponding energy bandpass; the peak burst luminosity; the isotropic-equivalent γ -ray burst energy; the isotropic-equivalent luminosity in X-ray at 10 h post-burst; the beaming fraction, calculated from the jet collimation angle θ_j as $f_b = 1 - \cos(\theta_j)$; the γ -ray energy release corrected for beaming fraction, $f_b E_{\gamma, \text{iso}}$; the host galaxy luminosity; the host star-formation rate; the burst offset from its host, in physical units and referred to the scale length of the host galaxy's light profile; and the absolute magnitude limit on any associated supernova. Empty cells are unconstrained by the data at present; values without uncertainties are known to roughly 20% precision. The values of $E_{\gamma, \text{iso}}$ and $L_{\gamma, \text{peak}}$ have been increased by a factor of four, representing a mean correction from the BATSE sample for converting these observed fluences to the 25–2,000 keV band.

*D.B.F. et al. (manuscript in preparation); M. Gladders et al. (manuscript in preparation).

redshifts and peak luminosities to one other (Table 2) strongly suggests a common origin for these events.

We find that SHBs are distinctly weaker explosions than long-duration GRBs. The lower energies that we infer from afterglow observations of GRB 050709 and GRB 050724 are consistent with the merger of a compact object binary, as seen in numerical simulations. Moreover, the two best-studied events show strong collimation; thus the true rate of SHBs is 30 to 100 times the observed rate.

The long-duration GRBs have been definitively associated with the deaths of massive stars^{32–35}. Two properties of the known SHBs argue against such an association: first, the fact that GRB 050724 occurred in an elliptical galaxy without active star formation²¹, and that GRB 050509B probably occurred in an elliptical as well⁵; and second, that GRB 050509B^{4,36} and GRB 050709 lack associated supernovae, a common feature of $z < 1$ long-duration GRBs, to very deep limits. Separately, the presence of SHBs among old stellar populations in elliptical galaxies argues against a magnetar origin, or a form of short-lived compact binary.

The locations of the SHBs with respect to their host galaxies are compatible with the kicks delivered to neutron-star binary systems at birth³⁷. Their occurrence in both star-forming (late-type) and non-star-forming (early-type) galaxies suggests that there may be a substantial range of lifetimes for the progenitor systems, perhaps extending to many gigayears.

In all respects, the emerging picture of SHB properties is consistent with an origin in the coalescence events of neutron star–neutron star or neutron star–black hole binary systems. The stage is now set for detailed studies of these exotic cosmic explosions, the most exciting of which would be the detection of their associated bursts of gravitational waves.

Received 11 August; accepted 30 August 2005.

- Zhang, B. & Meszaros, P. Gamma-ray bursts: progress, problems & prospects. *Int. J. Mod. Phys. A* **19**, 2385–2472 (2004).
- Kouveliotou, C. *et al.* Identification of two classes of gamma-ray bursts. *Astrophys. J.* **413**, L101–L104 (1993).
- Hurley, K. *et al.* Afterglow upper limits for four short-duration, hard spectrum gamma-ray bursts. *Astrophys. J.* **567**, 447–453 (2002).
- Kulkarni, S. R. *et al.* Supernova and low-velocity ejecta constraints for the short-hand γ -ray burst GRB 050509B. *Nature* (submitted).
- Bloom, J. S. *et al.* Closing in on a short-hard burst progenitor: constraints from early-time optical imaging and spectroscopy of a possible host galaxy of GRB 050509B. *Astrophys. J.* (submitted).
- Gehrels, N. *et al.* A short γ -ray burst apparently associated with an elliptical galaxy at redshift $z = 0.225$. *Nature* doi:10.1038/nature04142 (this issue).
- Schmidt, M. Luminosities and space densities of short gamma-ray bursts. *Astrophys. J.* **559**, L79–L82 (2001).
- Eichler, D., Livio, M., Piran, T. & Schramm, D. N. Nucleosynthesis, neutrino bursts and gamma-rays from coalescing neutron stars. *Nature* **340**, 126–128 (1989).
- Guetta, D. & Piran, T. The luminosity and redshift distributions of short-duration GRBs. *Astron. Astrophys.* **435**, 421–426 (2005).
- Villasenor, J. *et al.* Discovery of the short γ -ray burst GRB 050709. *Nature* doi:10.1038/nature04213 (this issue).
- Garmire, G. P., Bautz, M. W., Ford, P. G., Nousek, J. A. & Ricker, G. R. in *X-Ray and Gamma-Ray Telescopes and Instruments for Astronomy* (eds Truemper, J. E. & Tananbaum, H. D.) *Proc. SPIE* **4851**, 28–44 (2003).
- Condon, J. J. *et al.* The NRAO VLA Sky Survey. *Astron. J.* **115**, 1693–1716 (1998).
- Fox, D. B., Frail, D. A., Cameron, P. B. & Cenko, S. B. GRB050709: candidate X-ray afterglow. *GRB Circ. Netw.* **3585** (2004).
- Burrows, D. N. *et al.* Bright X-ray flares in gamma-ray burst afterglows. *Science* **150**, 1–5 (2005).
- Jensen, B. L. *et al.* GRB 050709: Chandra source optical counterpart. *GRB Circ. Netw.* **3589** (2005).
- Hjorth, J. *et al.* The optical afterglow of the short γ -ray burst GRB 050709. *Nature* doi:10.1038/nature04174 (this issue).
- Sirianni, M. *et al.* The photometric performance and calibration of the HST advanced camera for surveys. *Publ. Astron. Soc. Pacif.* (in the press); preprint at (<http://arxiv.org/astro-ph/0507614>) (2005).
- Bloom, J. S., Kulkarni, S. R. & Djorgovski, S. G. The observed offset distribution of gamma-ray bursts from their host galaxies: A robust clue to the nature of the progenitors. *Astron. J.* **123**, 1111–1148 (2002).
- Frail, D. A. *et al.* Beaming in gamma-ray bursts: evidence for a standard energy reservoir. *Astrophys. J.* **562**, L55–L58 (2001).
- Rhoads, J. E. The dynamics and light curves of beamed gamma-ray burst afterglows. *Astrophys. J.* **525**, 737–749 (1999).
- Berger, E. *et al.* A merger origin for short gamma-ray bursts inferred from the afterglow and host galaxy of GRB 050724. *Nature* (submitted); preprint at (<http://arxiv.org/astro-ph/0508115>) (2005).
- Kumar, P. The distribution of burst energy and shock parameters for gamma-ray bursts. *Astrophys. J.* **538**, L125–L128 (2000).
- Berger, E., Kulkarni, S. R. & Frail, D. A. A standard kinetic energy reservoir in gamma-ray burst afterglows. *Astrophys. J.* **590**, 379–385 (2003).
- Janka, H.-T. & Ruffert, M. *Stellar Collisions, Mergers and their Consequences* 333–358 (ASP Conf. Ser. 263, Astronomical Society of the Pacific, San Francisco, 2002).
- Li, L. & Paczyński, B. Transient events from neutron star mergers. *Astrophys. J.* **507**, L59–L62 (1998).
- Totani, T. Probing the cosmic star formation history by the brightness distribution of gamma-ray bursts. *Astrophys. J.* **511**, 41–55 (1999).
- Balázs, L. G., Bagoly, Z., Horváth, I., Mészáros, A. & Mészáros, P. On the difference between the short and long gamma-ray bursts. *Astron. Astrophys.* **401**, 129–140 (2003).
- Katz, J. I. & Canel, L. M. The long and the short of gamma-ray bursts. *Astrophys. J.* **471**, 915–920 (1996).
- Piran, T. *Compact Stars in Binaries* 489–500 (IAU Symp. 165, Kluwer Academic, Dordrecht, 1996).
- Eisenstein, D. J., Hogg, D. W. & Padmanabhan, N. GRB050509b, SDSS pre-burst observations. *GRB Circ. Netw.* **3418** (2005).
- Mannucci, F. *et al.* The supernova rate per unit mass. *Astron. Astrophys.* **433**, 807–814 (2005).
- Kulkarni, S. R. *et al.* Radio emission from the unusual supernova 1998bw and its association with the gamma-ray burst of 25 April 1998. *Nature* **395**, 663–669 (1998).
- Galama, T. J. *et al.* An unusual supernova in the error box of the gamma-ray burst of 25 April 1998. *Nature* **395**, 670–672 (1998).
- Hjorth, J. *et al.* A very energetic supernova associated with the γ -ray burst of 29 March 2003. *Nature* **423**, 847–850 (2003).
- Stanek, K. Z. *et al.* Spectroscopic discovery of the supernova 2003dh associated with GRB 030329. *Astrophys. J.* **591**, L17–L20 (2003).
- Hjorth, J. *et al.* GRB 050509B: constraints on short gamma-ray burst models. *Astrophys. J. Lett.* **630**, 117–120 (2005).
- Narayan, R., Paczynski, B. & Piran, T. Gamma-ray bursts as the death throes of massive binary stars. *Astrophys. J.* **395**, L83–L86 (1992).
- Morgan, A. *et al.* GRB 050709: Swift UVOT and XRT observations. *CGN Circ.* **3577** (2005).
- Schlegel, D. J., Finkbeiner, D. P. & Davis, M. Maps of dust infrared emission for use in estimation of reddening and cosmic microwave background radiation foregrounds. *Astrophys. J.* **500**, 525–553 (1998).
- Motohara, K. *et al.* CISCO: cooled infrared spectrograph and camera for OHS on the Subaru telescope. *Proc. Astron. Soc. Jpn.* **54**, 315–325 (2002).
- Fruchter, A. S. & Hook, R. N. Drizzle: a method for the linear reconstruction of undersampled images. *Publ. Astron. Soc. Pacif.* **114**, 144–152 (2002).
- Covino, S. *et al.* GRB 050724: a short-burst detected by Swift. *GCN Circ.* **3665** (2005).
- Peng, C. Y., Ho, L. C., Impey, C. D. & Rix, H.-W. Detailed structural decomposition of galaxy images. *Astron. J.* **124**, 266–293 (2002).
- Berger, E. *et al.* A common origin for cosmic explosions inferred from calorimetry of GRB030329. *Nature* **426**, 154–157 (2003).
- Friedman, A. S. & Bloom, J. S. Toward a more standardized candle using gamma-ray burst energetics and spectra. *Astrophys. J.* **627**, 1–25 (2005).
- Berger, E. *et al.* The afterglows, redshifts, and properties of Swift gamma-ray bursts. *Astrophys. J.* (submitted); preprint at (<http://arxiv.org/astro-ph/0505107>) (2005).

Acknowledgements Our GRB research is supported in part by funds from NSF, NASA, the Australian Research Council, and the Ministry of Education, Science, Culture, Sports, and Technology in Japan. The VLA is operated by the National Radio Astronomy Observatory, a facility of the NSF operated under cooperative agreement by Associated Universities, Inc. The Gemini Observatory is operated by the Association of Universities for Research in Astronomy (AURA), Inc., under a cooperative agreement with the NSF on behalf of the Gemini partnership. This work is based in part on data collected at Subaru Telescope, which is operated by the National Astronomical Observatory of Japan.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.A.F. (dfrail@nrao.edu) or D.B.F. (dfbx@astro.psu.edu).

A short γ -ray burst apparently associated with an elliptical galaxy at redshift $z = 0.225$

N. Gehrels¹, C. L. Sarazin², P. T. O'Brien³, B. Zhang⁴, L. Barbier¹, S. D. Barthelmy¹, A. Blustin⁵, D. N. Burrows⁶, J. Cannizzo^{1,7}, J. R. Cummings^{1,8}, M. Goad³, S. T. Holland^{1,9}, C. P. Hurkett³, J. A. Kennea⁶, A. Levan³, C. B. Markwardt^{1,10}, K. O. Mason⁵, P. Meszaros⁶, M. Page⁵, D. M. Palmer¹¹, E. Rol³, T. Sakamoto^{1,8}, R. Willingale³, L. Angelini^{1,7}, A. Beardmore³, P. T. Boyd^{1,7}, A. Breeveld⁵, S. Campana¹², M. M. Chester⁶, G. Chincarini^{12,13}, L. R. Cominsky¹⁴, G. Cusumano¹⁵, M. de Pasquale⁵, E. E. Fenimore¹¹, P. Giommi¹⁶, C. Gronwall⁶, D. Grupe⁶, J. E. Hill⁶, D. Hinshaw^{1,17}, J. Hjorth¹⁸, D. Hullinger^{1,10}, K. C. Hurley¹⁹, S. Klose²⁰, S. Kobayashi⁶, C. Kouveliotou²¹, H. A. Krimm^{1,9}, V. Mangano¹², F. E. Marshall¹, K. McGowan⁵, A. Moretti¹², R. F. Mushotzky¹, K. Nakazawa²², J. P. Norris¹, J. A. Nousek⁶, J. P. Osborne³, K. Page³, A. M. Parsons¹, S. Patel²³, M. Perri¹⁶, T. Poole⁵, P. Romano¹², P. W. A. Roming⁶, S. Rosen⁵, G. Sato²², P. Schady⁵, A. P. Smale²⁴, J. Sollerman²⁵, R. Starling²⁶, M. Still^{1,9}, M. Suzuki²⁷, G. Tagliaferri¹², T. Takahashi²², M. Tashiro²⁷, J. Tueller¹, A. A. Wells³, N. E. White¹ & R. A. M. J. Wijers²⁶

Gamma-ray bursts (GRBs) come in two classes¹: long (>2 s), soft-spectrum bursts and short, hard events. Most progress has been made on understanding the long GRBs, which are typically observed at high redshift ($z \approx 1$) and found in subluminal star-forming host galaxies. They are likely to be produced in core-collapse explosions of massive stars². In contrast, no short GRB had been accurately (<10'') and rapidly (minutes) located. Here we report the detection of the X-ray afterglow from—and the localization of—the short burst GRB 050509B. Its position on the sky is near a luminous, non-star-forming elliptical galaxy at a redshift of 0.225, which is the location one would expect^{3,4} if the origin of this GRB is through the merger of neutron-star or black-hole binaries. The X-ray afterglow was weak and faded below the detection limit within a few hours; no optical afterglow was detected to stringent limits, explaining the past difficulty in localizing short GRBs.

The new observations are from the Swift⁵ satellite, which features the hard X-ray wide-field Burst Alert Telescope (BAT), and rapid spacecraft slewing to point the narrow-field X-ray Telescope (XRT) and the Ultraviolet-optical Telescope (UVOT) at the burst. On 9 May 2005 at 04:00:19.23 UT, the BAT triggered and located GRB 050509B on board⁶. The BAT location is shown in Fig. 1 (large red circle) and the light curves in Fig. 2. The event is a single short spike with duration of 40 ± 4 ms. The burst has a ratio of 50–100 keV to 25–25 keV fluences of 1.4 ± 0.5 , which is consistent with, but in the soft portion of, the short/hard population detected by the first extensive

GRB survey made with the Burst and Transient Source Experiment (BATSE). The 15–150 keV fluence is $(9.5 \pm 2.5) \times 10^{-9}$ erg cm⁻², which is the lowest imaged by BAT so far and is just below the short GRB fluence range detected by BATSE (adjusted for the different energy ranges of the two instruments).

Swift slewed promptly and XRT started acquiring data 62 s after the burst ($T+62$ s, where T is the BAT trigger time). Ground-processed data revealed an uncatalogued X-ray source near the centre of the BAT error circle containing 11 photons (5.7σ significance due to near-zero background in image) in the first 1,640 s of integration time. The XRT position is shown with respect to the Digitized Sky Survey (DSS) field in Fig. 1. A *Chandra* target-of-opportunity observation of the XRT error circle was performed on 11 May at 4:00 UT for 50 ks, with no sources detected in the XRT error circle. The light curve combining BAT, XRT and *Chandra* data are shown in Fig. 3. The UVOT observed the field starting at $T+60$ s. No new optical/ultraviolet sources were found in the XRT error circle to V-band magnitude > 19.7 for $t < 300$ min.

Swift has provided the first accurate localization of a short GRB. No optical afterglow was detected to stringent limits (R-band magnitude > 25 at 25 h; ref. 7). When the XRT error circle is plotted on the R-band image we obtained⁸ with the Very Large Telescope (VLT), several faint objects are seen in the error circle, some of which are extended and could be high-redshift galaxies^{9,10}. It is possible the burst occurred in one of these. However, the centre of the XRT error circle lies only 9.8'' away from the centre of the large E1 elliptical

¹NASA/Goddard Space Flight Center, Greenbelt, Maryland 20771, USA. ²Department of Astronomy, University of Virginia, Charlottesville, Virginia 22903-0818, USA.

³Department of Physics and Astronomy, University of Leicester, Leicester, LE1 7RH, UK. ⁴Department of Physics, University of Nevada, Las Vegas, Nevada 89154-4002, USA.

⁵Mullard Space Science Laboratory, University College London, Dorking RH5 6NT, UK. ⁶Department of Astronomy and Astrophysics, Penn State University, University Park, Pennsylvania 16802, USA. ⁷Joint Center for Astrophysics, University of Maryland, Baltimore County, Baltimore, Maryland 21250, USA. ⁸National Research Council, 2101 Constitution Ave NW, Washington DC 20418, USA. ⁹Universities Space Research Association, 10211 Wincopin Circle, Suite 500, Columbia, Maryland 21044-3432, USA.

¹⁰Department of Astronomy, University of Maryland, College Park, Maryland 20742, USA. ¹¹Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA.

¹²INAF—Osservatorio Astronomico di Brera, Via Biancamano 46, I-23807 Merate, Italy. ¹³Università degli studi di Milano Bicocca, Piazza delle Scienze 3, I-20126 Milano, Italy.

¹⁴Department of Physics and Astronomy, Sonoma State University, Rohnert Park, California 94928, USA. ¹⁵INAF—Istituto di Astrofisica Spaziale e Cosmica, Via Ugo La Malfa 153, I-90146 Palermo, Italy. ¹⁶ASI Science Data Center, Via Galileo Galilei, I-00044 Frascati, Italy. ¹⁷SP Systems Inc., 7500 Greenway Center Drive, Greenbelt, Maryland 20770, USA. ¹⁸Niels Bohr Institute, University of Copenhagen, DK-2100 Copenhagen, Denmark. ¹⁹UC Berkeley Space Sciences Laboratory, Berkeley, California 94720-7450, USA.

²⁰Thüringer Landessternwarte Tautenburg, Sternwarte 5, D-07778 Tautenburg, Germany. ²¹NASA/Marshall Space Flight Center, NSSTC, XD-12, 320 Sparkman Drive, Huntsville, Alabama 35805, USA. ²²Institute of Space and Astronautical Science, JAXA, Kanagawa 229-8510, Japan. ²³Universities Space Research Association, NSSTC, XD-12, 320 Sparkman Drive, Huntsville, Alabama 35805, USA. ²⁴Office of Space Science, NASA Headquarters, Washington DC 20546, USA. ²⁵Stockholm Observatory, Department of Astronomy, AlbaNova, 106 91 Stockholm, Sweden. ²⁶Astronomical Institute "Anton Pannekoek", University of Amsterdam, Kruislaan 403, 1098 SJ Amsterdam, The Netherlands.

²⁷Department of Physics, Saitama University, Sakura, Saitama 338-8570, Japan.

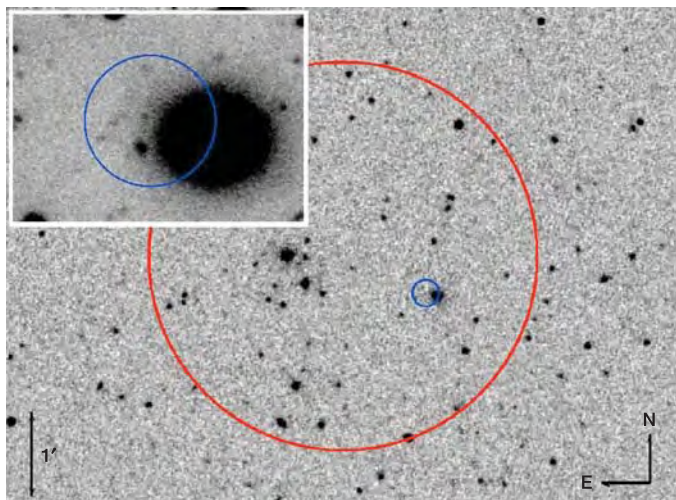


Figure 1 | Optical images of the region of GRB 050509B showing the association with a large elliptical galaxy. The Digitized Sky Survey image. The large red circle is the BAT position error circle, and the smaller blue circle is the XRT position error circle. The BAT position is 12 h 36 m 18 s, +28° 59' 28" (J2000) with a 2.3' error radius (90% containment). The XRT, operating in its most sensitive 'photon counting' mode, derived a position of 12 h 36 m 13.58 s, +28° 59' 01.3" (J2000), with a positional accuracy of 9.3" (90% containment radius; larger than the typical XRT 4" accuracy, owing to weakness of burst). This position takes into account the low counting statistics, cluster emission in the field and astrometric corrections¹⁰ to the 2MASS coordinate system. Many of the extended objects are likely to be galaxies in the cluster NSC J123610+28590131. The inset shows a blow-up of the region of the XRT error circle from an R-band image obtained⁸ using FORS2 on the 8.2-m VLT-Antu telescope at the European Southern Observatory/Paranal on 11 May UT, 1.85 days after the burst. The extended source to the right (west) is the luminous elliptical galaxy 2MASX J12361286+285858026, which we suggest to be the likely host of the burst. Other objects in the error circle are not identified, but appear to be faint galaxies either associated with the same cluster as the elliptical galaxy or at higher redshift. The VLT image consists of fifteen 3-min frames taken under good conditions ($\sim 1''$ seeing).

galaxy 2MASX J12361286+2858580 (ref. 10) at a redshift of 0.225 (ref. 11), which is located in the cluster NSC J123610+285901 (refs 12, 13). This is a luminous giant elliptical galaxy; its 2 Micron All Sky Survey (2MASS) magnitude of $K = 14.1$ corresponds to a luminosity of $4 \times 10^{11} L_{\odot} \approx 3L^*$, where $L_{\odot}$ is the luminosity of the Sun and L^* is the luminosity of a typical galaxy, assuming standard cosmology. Our *Chandra* image shows that this is the central dominant galaxy in one of two merging subclusters in this bimodal cluster. Although caution is always prudent for a posteriori statistics, the association with this galaxy seems unlikely to be coincidental. The probability of a random location being within $10''$ of a galaxy with an apparent magnitude at least this bright is $\sim 10^{-3}$. Moreover, galaxies this luminous are relatively rare; the comoving number density¹⁴ of galaxies at least this luminous is $\sim 5 \times 10^{-5} \text{ Mpc}^{-3}$; the probability of lying within $10''$ of a randomly located one at $z \leq 0.225$ is $\sim 10^{-4}$. Note that this is the first GRB of ~ 80 with accurate optical localizations to be near a bright elliptical on the sky.

The likely association between GRB 050509B and 2MASX J12361286+2858580 is difficult to understand if the GRB resulted from any mechanism involving recent star formation. The galaxy type for the suggested host galaxy is very different from those found for long GRBs; their hosts are typically subluminescent and blue¹⁵ and show strong emission lines associated with star formation¹⁶. As is true of most giant ellipticals in clusters, 2MASX J12361286+2858580 has no indications of ultraviolet or optical line emission¹⁰. Our UVOT images clearly detect the galaxy in the optical, but not in the

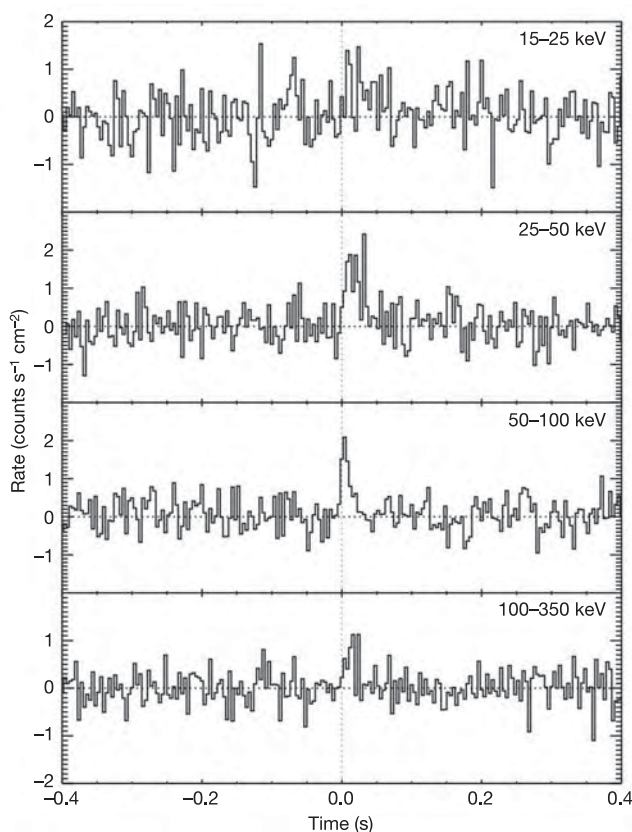


Figure 2 | BAT light curves for the short GRB 050509B, showing the short duration of this GRB. The light curves are given in four photon energy bands with the band identified in the upper right of each panel. The peak has a duration of 40 ± 4 ms (90% containment of counts). There is no detectable emission except from $T-30$ ms to $T+30$ ms, confirming the 'short' aspect of this burst. The successful trigger criterion for the GRB was in the 25–100 keV band. The peak count rate measured by BAT is $\sim 2,100 \text{ counts s}^{-1}$ in the 15–150 keV band at $T+5$ ms. The BAT data (40 ms of data centred on $T+23$ ms) are well fitted by a simple power-law model with a photon index of 1.5 ± 0.4 , a normalization at 50 keV of $(2.0 \pm 0.5) \times 10^{-2} \text{ photons cm}^{-2} \text{ s}^{-1} \text{ keV}^{-1}$ and a peak flux of $2.53 \pm 0.33 \text{ photons cm}^{-2} \text{ s}^{-1}$ (all in 15–150 keV and 90% confidence level).

ultraviolet (UVM2 220-nm and UVW2 188-nm filters), as expected for an elliptical galaxy—implying little or no contribution from young, hot stars. The 3σ upper limit at 188 nm gives a limit to the star-formation rate¹⁷ of $< 0.2 M_{\odot} \text{ yr}^{-1}$, where $M_{\odot}$ is the mass of the Sun. It is improbable that we will find a massive-star core collapse or young magnetar in this galaxy. In addition, the isotropic energy of $1.1 \times 10^{48} \text{ erg}$ (15–150 keV, $z = 0.225$, where the k -correction factor is typically 1 to 10) is $> 10^2$ times higher than that of the 27 December 2004 giant flare from SGR 1806–20 (refs 18, 19). Thus, it is unlikely that this burst was an SGR-type flare.

On the other hand, 2MASX J12361286+2858580 is a very propitious site for a neutron star–neutron star or neutron star–black hole merger. As *Chandra* observations have shown²⁰, giant ellipticals, especially those dominant in their cluster, have large populations of low-mass X-ray binaries containing accreting neutron stars and black holes. Further, a high fraction ($\geq 50\%$) of the low-mass X-ray binaries in ellipticals are located in globular clusters²¹ because close binary systems containing at least one compact object can easily be formed dynamically in globular clusters. Although there is less direct evidence that close neutron star–neutron star binaries can form easily in globular clusters, the double-neutron-star system PSR B2127+11C in the Galactic globular cluster M15 is an example of such a binary²², and has a merger lifetime of $\sim 2 \times 10^8 \text{ yr}$. In fact, of

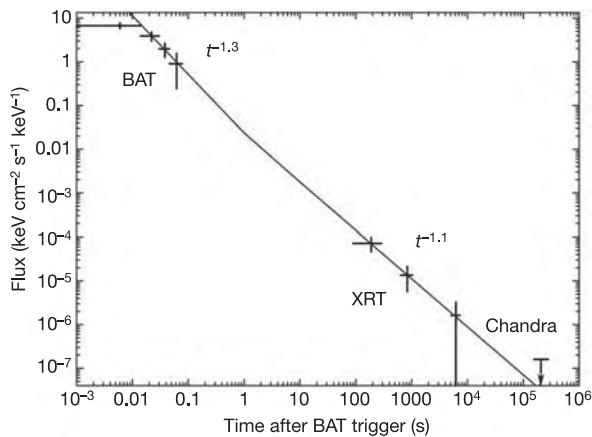


Figure 3 | X-ray afterglow light curve for GRB 050509B showing weak flux falling off to undetectability after 10^4 s. The decline in the 1-keV X-ray emission from GRB 050509B from BAT, XRT and Chandra data is shown. The BAT points were calculated by extrapolating the BAT spectrum with measured photon index Γ (flux = constant $\times E^{-\Gamma}$) of 1.5 down to 1 keV. The XRT fluxes are unabsorbed values derived by fixing the photon index to 1.5 (that is, that found by the BAT) and using the Galactic column density of $1.5 \times 10^{20} \text{ cm}^{-2}$. The initial flux is $(3.57 \pm 1.08) \times 10^{-13} \text{ erg cm}^{-2} \text{ s}^{-1}$ between 0.3 and 10 keV. The XRT time decay index α (flux = constant $\times \text{time}^{-\alpha}$) is 1.10 (0.57 to 2.36; 90% confidence) and the BAT index is 1.34 (0.27 to 2.87; 90% confidence). The BAT index is poorly constrained because the burst is so short. The BAT and XRT data are also consistent at the 90% confidence level with a single decay index of 1.20 (1.12 to 1.29). The Chandra 99% confidence flux upper limit lies above the decay curves. The error bars in the figure are at the 1σ 68% confidence level.

all galaxy types, large ellipticals (particularly cluster dominants) are the most likely place to find double compact binary systems, owing to their large populations of globular clusters.

The centre of the XRT error circle is in the outer regions of the elliptical galaxy, although the circle extends nearly to the galaxy centre. A location at large radius would be consistent with the binary merger model for short GRBs²³. Neutron star–neutron star binaries often obtain significant kick velocities ($100\text{--}1,000 \text{ km s}^{-1}$) from the supernova that creates each neutron star²⁴. If we ignore the effects of the galactic potential, a neutron star–neutron star binary moving at $1,000 \text{ km s}^{-1}$ would travel 100 kpc in 10^8 yr . The projected distance of the centre of the XRT position from the centre of the galaxy is about 35 kpc; the range over the error circle is about 2–70 kpc. Thus, the neutron star–neutron star binary might have reached this distance before merging, even if it started from a more central location. Alternatively, the binary may have formed in a globular cluster; globular clusters have a broader radial distribution than field stars in ellipticals, which could explain the large projected radius of the GRB.

The X-ray emission for GRB 050509B is faint, being the weakest afterglow of any of the 15 GRBs that XRT has promptly observed (a factor of ~ 200 weaker than the XRT average). For BATSE bursts, studies were done of the post-burst emission by summing large numbers of GRB lightcurves^{25,26}. The post-burst emission was found to be weaker for short bursts than for long events, consistent with the GRB 050509B. For typical shock parameters, the early X-ray afterglow emission is probably below the cooling frequency²⁷; in this regime, the weak afterglow is consistent with the low-density medium around an evolved compact binary progenitor. A more critical factor to define the low X-ray flux may be the small energy injection involved, as the prompt emission for GRB 050509B is also the weakest of the BAT GRBs. If the redshift is 0.225, then the afterglow is >100 times less luminous than that of typical long-burst afterglows and the isotropic energy is $\sim 10^{-4}$ that of typical

long GRBs (about the same as the lowest-luminosity, unusual GRB 980425).

Before Swift's observations, it was predicted²⁸ that short GRBs would have faint optical afterglows, particularly so if they occurred in low-density regions like those around evolved stars. This prediction is consistent with the lack of optical detection to stringent limits for GRB 050509B, although we bear in mind that this burst is weak compared to other short GRBs. It is likely that the X-ray afterglow will remain a key to understanding short bursts.

The X-ray afterglow from this short GRB can constrain outflow parameters. The fact that the X-rays are fading as early as 62 s puts a limit on the initial Lorentz factor of $\Gamma_0 \geq 70 n_{-2}^{-1/8} E^{1/8}$ for $z = 0.225$ (n_{-2} is ambient density in units of 10^{-2} cm^{-3} , and E is the isotropic energy in units of 10^{48} erg), showing that short GRBs are highly relativistic events.

Another interesting aspect of the localization of GRB 050509B is that the burst is faint and yet has a bright galaxy in its error circle. There are five previous short GRBs with fluences one to three orders of magnitude larger than GRB 050509B that have had their arcminute-sized error boxes searched for bright galaxies^{29,30}. There are galaxies in each error box of brightness comparable to or less than 2MASX J12361286+2858580, but none much brighter—as one might expect for these brighter GRBs. This does not contradict a merger model for short GRBs, because, although giant elliptical galaxies are a rich environment for mergers, most would occur in the more numerous, fainter, star-forming galaxies. Thus star-forming galaxies harbour both massive stars and evolved binaries, whereas ellipticals have almost no star formation and are highly deficient in short-lived massive stars. The detection of GRB 050509B near an elliptical galaxy is an important observation for short bursts because the association with a large elliptical galaxy is evidence against a collapsar origin, whereas an association with a star-forming galaxy would have left the question unanswered. There may be more than one origin of short GRBs, but this particular short event has a high probability of being unrelated to star formation and of being caused by a binary merger.

Received 31 May; accepted 10 August 2005.

- Kouveliotou, C. *et al.* Identification of two classes of gamma-ray bursts. *Astrophys. J.* **413**, L101–L104 (1993).
- van Paradijs, J., Kouveliotou, C. & Wijers, R. A. M. J. Gamma-ray burst afterglows. *Annu. Rev. Astron. Astrophys.* **38**, 379–425 (2000).
- Eichler, D., Livio, M., Piran, T. & Schramm, D. N. Nucleosynthesis, neutrino bursts and gamma-rays from coalescing neutron stars. *Nature* **340**, 126–127 (1989).
- Mochkovitch, R., Hernanz, M., Isern, J. & Martin, X. Gamma-ray bursts as collimated jets from neutron star/black hole mergers. *Nature* **361**, 236–237 (1993).
- Gehrels, N. *et al.* The Swift gamma ray burst mission. *Astrophys. J.* **611**, 1005–1020 (2004).
- Hurkett, C. *et al.* Swift detection of GRB050509B: A short duration burst. *GCN Circ.* **3381** (2005).
- Cenko, S. B. *et al.* GRB050509B: No optical variability in XRT error circle. *GCN Circ.* **3409** (2005).
- Hjorth, J. *et al.* GRB050509B: Optical observations with the VLT. *GCN Circ.* **3410** (2005).
- Cenko, S. B. *et al.* GRB050509B: Further analysis of Keck LRIS images. *GCN Circ.* **3401** (2005).
- Bloom, J. S. *et al.* Closing in on a short-hard burst progenitor: constraints from early-time optical imaging and spectroscopy of a possible host galaxy of GRB050509B. *Astrophys. J.* (submitted).
- Prochaska, J. X. *et al.* Keck/DEIMOS spectrum of possible host galaxy for GRB050509B. *GCN Circ.* **3390** (2005).
- Kennea, J. A. *et al.* GRB050509B: Swift XRT Position. *GCN Circ.* **3383** (2005).
- Gal, R. R. *et al.* The Northern Sky Optical Cluster Survey. II. An objective cluster catalog for 5800 square degrees. *Astron. J.* **125**, 2064–2084 (2003).
- Kochanek, C. S. *et al.* The K-band galaxy luminosity function. *Astrophys. J.* **560**, 566–579 (2001).
- Le Floc'h, E. *et al.* Are the hosts of gamma-ray bursts sub-luminous and blue galaxies? *Astron. Astrophys.* **400**, 499–510 (2003).
- Vreeswijk, P. M. *et al.* VLT Spectroscopy of GRB 990510 and GRB 990712: Probing the faint and bright ends of the gamma-ray burst host galaxy population. *Astrophys. J.* **546**, 672–680 (2001).

17. Madau, P., Pozzetti, L. & Dickinson, M. The star formation history of field galaxies. *Astrophys. J.* **498**, 106–116 (1998).
18. Hurley, K. *et al.* An exceptionally bright flare from SGR 1806–20 and the origin of short duration gamma-ray bursts. *Nature* **434**, 1098–1103 (2005).
19. Palmer, D. M. *et al.* A giant gamma-ray flare from the magnetar SGR 1806–20. *Nature* **434**, 1107–1109 (2005).
20. Sarazin, C. L., Irwin, J. A. & Bregman, J. N. Resolving the mystery of X-ray-faint elliptical galaxies: Chandra X-ray observations of NGC 4697. *Astrophys. J.* **544**, L101–L105 (2000).
21. Sarazin, C. L. *et al.* Low-mass X-ray binaries and globular clusters in early-type galaxies. *Astrophys. J.* **595**, 743–759 (2003).
22. Anderson, S. B. *et al.* Discovery of two radio pulsars in the globular cluster M15. *Nature* **346**, 42–44 (1990).
23. Bloom, J. S., Sigurdsson, S. & Pols, O. R. The spatial distribution of coalescing neutron star binaries; implications for gamma-ray bursts. *Mon. Not. R. Astron. Soc.* **305**, 763–769 (1999).
24. Willems, B., Kalogera, V. & Henninger, M. Pulsar kicks and spin tilts in the close double neutron stars PSR J0737–3039, PSR B1534 + 12, and PSR B1913 + 16. *Astrophys. J.* **616**, 414–438 (2004).
25. Lazatti, D., Ramirez-Ruiz, E. & Ghisellini, G. Possible detection of hard X-ray afterglows of short gamma-ray bursts. *Astron. Astrophys.* **379**, L39–L43 (2001).
26. Connaughton, V. BATSE Observations of gamma-ray burst tails. *Astrophys. J.* **567**, 1028–1036 (2002).
27. Panaitescu, A. & Kumar, P. Analytic light curves of gamma-ray burst afterglows: homogeneous versus wind external media. *Astrophys. J.* **543**, 66–76 (2000).
28. Panaitescu, A., Kumar, P. & Narayan, R. Observational prospects for afterglows of short-duration gamma-ray bursts. *Astrophys. J.* **561**, L171–L174 (2001).
29. Nakar, E. *et al.* The distance of short-hard GRBs and the SGR connection. Preprint at (<http://arXiv.org/astro-ph/0502148>) (2005).
30. Hurley, K. *et al.* Afterglow upper limits for four short-duration, hard spectrum gamma-ray bursts. *Astrophys. J.* **567**, 447–453 (2002).

Acknowledgements The authors acknowledge support from ASI, NASA and PPARC.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to N.G. (gehrels@milkyway.gsfc.nasa.gov).

Discovery of the short γ -ray burst GRB 050709

J. S. Villaseñor¹, D. Q. Lamb², G. R. Ricker¹, J.-L. Atteia³, N. Kawai⁴, N. Butler¹, Y. Nakagawa⁵, J. G. Jernigan⁶, M. Boer⁷, G. B. Crew¹, T. Q. Donaghy², J. Doty⁸, E. E. Fenimore⁹, M. Galassi⁹, C. Graziani², K. Hurley⁶, A. Levine¹, F. Martel¹⁰, M. Matsuoka¹¹, J.-F. Olive⁷, G. Prigozhin¹, T. Sakamoto¹², Y. Shirasaki¹³, M. Suzuki¹⁴, T. Tamagawa⁵, R. Vanderspek¹, S. E. Woosley¹⁵, A. Yoshida⁶, J. Braga¹⁶, R. Manchanda¹⁷, G. Pizzichini¹⁸, K. Takagishi¹⁹ & M. Yamauchi¹⁹

Gamma-ray bursts (GRBs) fall into two classes: short-hard and long-soft bursts^{1–3}. The latter are now known to have X-ray⁴ and optical⁵ afterglows, to occur at cosmological distances⁶ in star-forming galaxies⁷, and to be associated with the explosion of massive stars^{8,9}. In contrast, the distance scale, the energy scale and the progenitors of the short bursts have remained a mystery. Here we report the discovery of a short-hard burst whose accurate localization has led to follow-up observations that have identified the X-ray afterglow¹⁰ and (for the first time) the optical afterglow^{10,11} of a short-hard burst; this in turn led to the identification of the host galaxy of the burst as a late-type galaxy at $z = 0.16$ (ref. 10). These results show that at least some short-hard bursts occur at cosmological distances in the outskirts of galaxies, and are likely to be caused by the merging of compact binaries.

On 9 July 2005, at 22:36:37 UT, the Soft X-Ray Camera (SXC), the Wide-Field X-Ray Monitor (WXM) and the French Gamma Telescope (FREGATE) instruments on board the High Energy Transient Explorer 2 satellite (HETE¹²) detected GRB 050709, a short-hard pulse followed by a long-soft bump from the same location. Figure 1 shows the WXM and SXC localizations for the burst, and the location of the X-ray and optical afterglows. Figures 2 and 3 show the time history of the entire burst and of the short-hard pulse in several energy bands.

Figure 4 compares the best-fit spectral model and the spectral data for the short-hard pulse. Table 1 gives the best-fit spectral parameters for different time intervals during the burst. The spectrum of the short pulse is hard and that of the intense first peak of the pulse (corresponding to the first 0.2 s of the burst) is even harder.

The duration and the peak energy $E_{\text{peak}}^{\text{obs}}$ of the spectrum of the short-hard pulse are consistent with those of short-hard GRBs^{13,14}. We note that its duration is much shorter and its spectrum is much harder than these were for GRB 020531, the other short-hard burst localized by HETE¹⁵. We also note that the time history and the spectral properties of GRB 050709 are similar to those of several Burst and Transient Source Experiment (BATSE) bursts, including GRBs 921022, 990516 and 990712 (refs 16, 17 and see Table 1). Table 2 gives the emission properties of the burst. The gamma-ray to X-ray fluence ratio of the short-hard pulse is 3.1, which is also consistent with those of BATSE short-hard bursts.

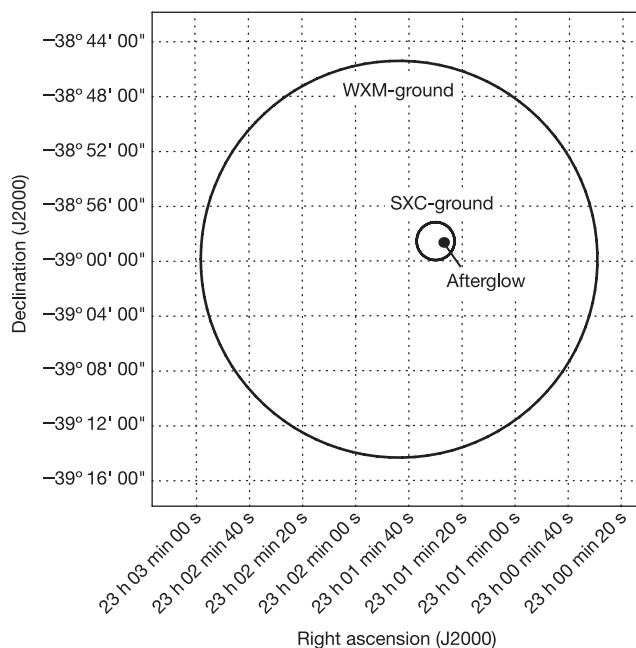


Figure 1 | Sky map showing the HETE localization error circles for GRB 050709 and the location of the X-ray and optical afterglow. The WXM obtained a localization in flight. However, the spacecraft attitude-control system was not locked at the time of the trigger, resulting in a drift of the satellite pointing direction, and real-time aspect was not available. Consequently, the location was not distributed in real time. Ground analysis of the data from the optical cameras provided reliable spacecraft aspect information, despite the spacecraft drift rate. A GCN Notice was sent out at 22:00:09 UT on 10 July 2005, after ground determination of the spacecraft aspect. Ground analysis of the WXM data produced a location with a 90% confidence region that is a circle centred at right ascension (RA) +23 h 01 min 44 s and declination (dec.) $-38^{\circ} 59' 52''$ (J2000) with a radius of $14.5'$ (large circle labelled 'WXM-ground'). Ground analysis of the SXC data yielded a refined location with a 90% confidence region that is a circle centred at RA = +23 h 01 min 30 s; dec. = $-38^{\circ} 58' 33''$ (J2000) with a radius of $1.34'$ (small circle labelled 'SXC-ground'). The location of the X-ray and optical afterglow is labelled 'afterglow'.

¹MIT Kavli Institute, Massachusetts Institute of Technology, 70 Vassar Street, Cambridge, Massachusetts 02139, USA. ²Department of Astronomy and Astrophysics, University of Chicago, 5640 South Ellis Avenue, Chicago, Illinois 60637, USA. ³Laboratoire d'Astrophysique, Observatoire Midi-Pyrénées, 14 Avenue E. Belin, 31400 Toulouse, France.

⁴Department of Physics, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo 152-8551, Japan. ⁵Department of Physics and Mathematics, Aoyama Gakuin University, Fuchinobe 5-10-1, Sagamihara, Kanagawa 229-8558, Japan. ⁶University of California at Berkeley, Space Sciences Laboratory, Berkeley, California 94720-7450, USA.

⁷Centre d'Etude Spatiale des Rayonnements, Observatoire Midi-Pyrénées, 9 Avenue de Colonel Roche, 31028 Toulouse cedex 4, France. ⁸Noqsi Aerospace Ltd, 2822 South Nova Road, Pine, Colorado 80470, USA. ⁹Los Alamos National Laboratory, PO Box 1663, Los Alamos, New Mexico 87545, USA. ¹⁰Espace Inc., 30 Lynn Avenue, Hull, Massachusetts 02045, USA. ¹¹Tsukuba Space Center, Japan Aerospace Exploration Agency, Tsukuba, Ibaraki 305-8505, Japan. ¹²NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA. ¹³National Astronomical Observatory, Osawa 2-21-1, Mitaka, Tokyo 181-8588, Japan. ¹⁴RIKEN (Institute of Physical and Chemical Research), 2-1 Hirosawa, Wako, Saitama 351-0198, Japan. ¹⁵Department of Astronomy and Astrophysics, University of California at Santa Cruz, 477 Clark Kerr Hall, Santa Cruz, California 95064, USA.

¹⁶Instituto Nacional de Pesquisas Espaciais, Avenida Dos Astronautas 1758, São José dos Campos 12227-010, Brazil. ¹⁷Department of Astronomy and Astrophysics, Tata Institute of Fundamental Research, Homi Bhabha Road, Mumbai, 400 005, India. ¹⁸INAF/IASF Bologna, Via Gobetti 101, 40129 Bologna, Italy. ¹⁹Faculty of Engineering, Miyazaki University, Gakuen Kibanadai Nishi, Miyazaki 889-2192, Japan.

The isotropic-equivalent energy E_{iso} of the short-hard pulse is a factor of $\sim 1,000$ smaller than is typical of long GRBs and implies that the energy E_γ radiated by the short-hard pulse in γ rays is at least 40 times less than is typical for long GRBs¹⁸. The very small value of E_{iso} also places the short-hard pulse off the $E_{\text{iso}}-E_{\text{peak}}$ relation found by refs 19 and 20 for long GRBs by a factor of $\sim 1,000$ in E_{iso} . The luminosity L_{iso} of the short-hard pulse is also very small, and places it off the $L_{\text{iso}}-E_{\text{peak}}$ correlation found by refs 20 and 21 for long GRBs by a factor of ~ 100 in L_{iso} . These three results strengthen the conclusion that GRB 050709 is not a long-soft burst.

There is evidence that many short-hard bursts observed by the BATSE and Konus instruments exhibit a long-soft bump following the initial hard pulse that is similar to that seen by HETE in GRB

050709 (refs 22–24). We have already mentioned the similarity of the time history and spectral properties of GRB 050709 and those of GRBs 921022, 990516 and 990712. References 22 and 23 reported evidence that BATSE and Konus short-hard bursts, respectively, are followed by a 30–200 s period of long-soft emission having a spectrum that is consistent with that of the long-soft bump in GRB 050709; ref. 24 reported evidence that BATSE short-hard bursts show an excess of soft emission from 20 s to 100–300 s after the burst.

The fluence of the long-soft bump is much greater than that of the short-hard pulse, unlike what is seen in soft gamma repeater (SGR) giant flares^{25,26}. In addition, we have analysed the time history of the long-soft bump and find no evidence for brightness oscillations of the kind that characterize the long-soft bump of SGR giant flares^{25,26} (see Fig. 2); however, we can place only a weak constraint on the amplitude of any such oscillations because the signal-to-noise ratio of the light curve of the long-soft bump is low. We have also searched for any evidence of stochastic variability of the long-soft bump and find none (see Fig. 2). Finally, we have searched for evidence of spectral evolution during the long-soft bump and find none (see Table 1).

The most natural interpretation of the long-soft bump is that it is the beginning of the afterglow. Its time history and spectrum are consistent with those expected for an afterglow, as is the lack of any time variability or spectral evolution. The ratio of the fluence in the short-hard pulse to that in the short-hard pulse plus the long-soft bump implies a radiative efficiency of $<25\%$ for the prompt phase. If

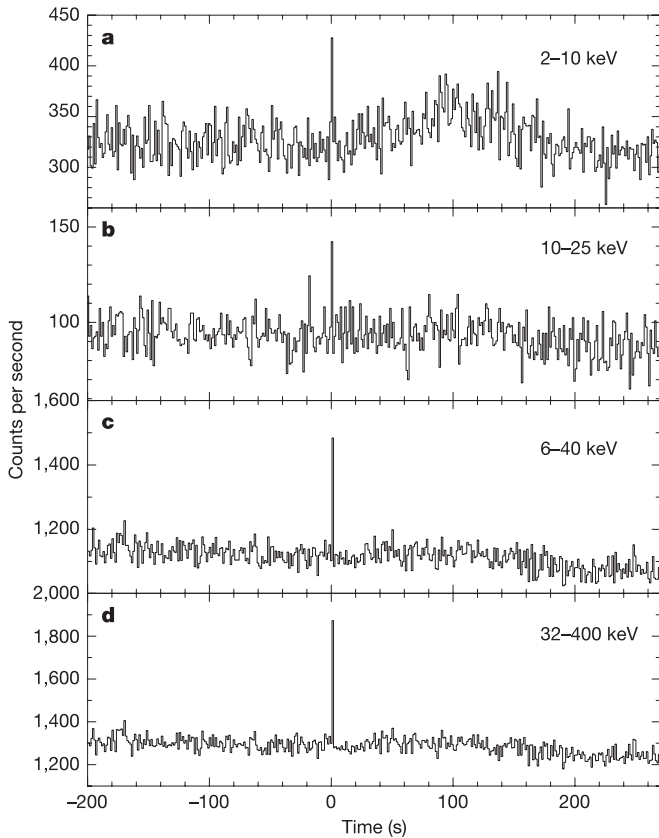


Figure 2 | Time history of GRB 050709. From top to bottom: Time history observed by WXM in the 2–10 keV energy band (a) and in the 2–25 keV energy band (b); time history observed by FREGATE in the 6–40 keV energy band (c) and in the 30–400 keV energy band (d). The event is a short-hard spike of duration $T_{90} = 220 \pm 50$ ms in the 2–25 keV energy band and 70 ± 10 ms in the 30–400 keV energy band, followed ~ 25 seconds later by a long-soft bump of duration $T_{90} = 130 \pm 7$ s in the 2–25 keV energy band. We have performed an FFT on the time history of the long-soft bump for the time interval 5–175 s after the trigger. We find no evidence for any coherent brightness oscillations in the period range 1–10 s, and derive 3σ upper limits on the amplitude of any such oscillations of 77% and 89% at periods of 1 and 5 s, respectively. We have also computed the variability measure V (ref. 31) for the long-soft bump, using smoothing timescales of 10, 20, 30 and 40 s. We find $V = 0.005 \pm 0.013$, 0.003 ± 0.026 , 0.007 ± 0.031 and 0.025 ± 0.036 , respectively. We therefore find no evidence for a non-zero value of V . If the long-soft bump is the afterglow and its peak at $t_{\text{peak}} \approx 100$ s corresponds to the time at which the fireball decelerates, a consistent solution exists in which $z = 0.16$; the GRB jet has an isotropic-equivalent kinetic energy of $E_{\text{KE}} = (1 - \cos\theta_{\text{jet}})E_{\text{iso}}^{\text{total}}/\eta \approx 5 \times 10^{49}$ erg, where $E_{\text{iso}}^{\text{total}} = E_{\text{iso}}(\text{pulse}) + E_{\text{iso}}(\text{bump}) \approx 1 \times 10^{50}$ erg, $\theta_{\text{jet}} \approx 0.3$ is the jet opening angle¹⁸, $\eta = 0.2$ is the radiative efficiency, the relativistic bulk $\Gamma \approx 100$, and the jet expands into a low-density medium having a number density $n \approx 10^{-2} \text{ cm}^{-3}$ (ref. 22).

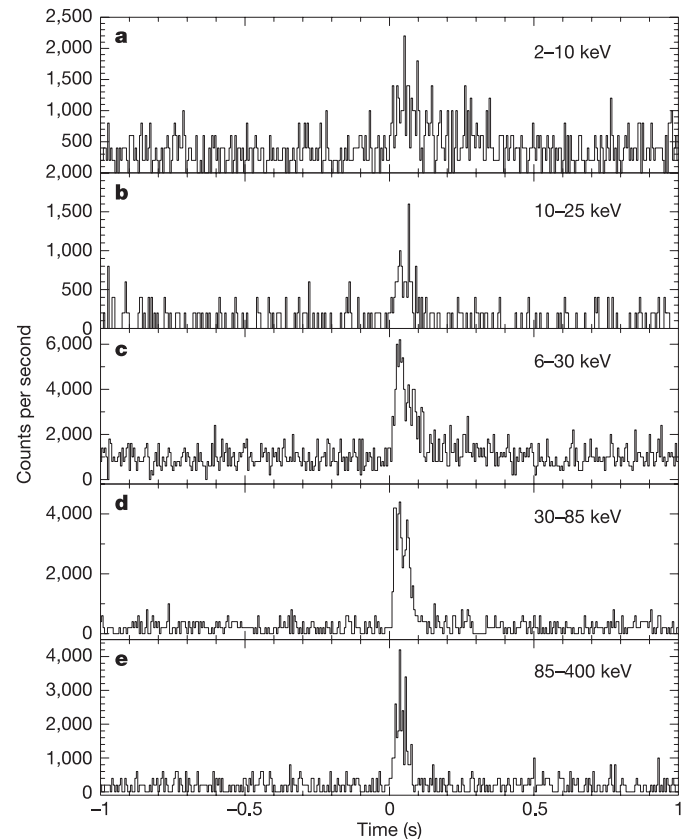


Figure 3 | Time history of the short-hard pulse of GRB 050709. Time histories observed by the WXM in the 2–10 keV energy band (a) and the 10–25 keV band (b); and by FREGATE in the 6–30 keV energy band (c), the 30–85 keV energy band (d), and the 85–400 keV energy band (e), plotted in 5 ms time bins. The pulse has a duration of $T_{90} = 220 \pm 50$ ms in the 2–25 keV energy band and 70 ± 10 ms in the 30–400 keV energy band, and exhibits no detectable emission before $T = 0$ or after $T = 400$ ms, confirming the short, hard nature of the pulse.

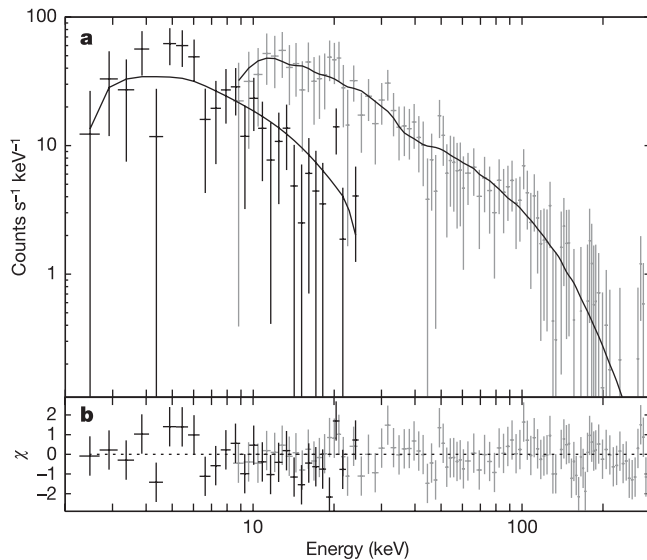


Figure 4 | Comparison of the observed count spectrum and the best-fit PLE model for the short-hard pulse of GRB 050709. **a**, Comparison of the counts in the WXM energy loss channels (lower energies) and the FREGATE energy loss channels (higher energies) and those predicted by the best-fit PLE model (smooth curves). Error bars are one sigma (68% confidence limits). **b**, Residuals to the fit. Error bars are one sigma. The short-hard pulse exhibits emission at all energies, confirming that its spectrum is hard.

the peak of the long-soft bump at ~ 100 s corresponds to the time at which the fireball decelerates, a consistent solution exists in which $z = 0.16$, the GRB jet has an isotropic-equivalent kinetic energy $E_{\text{KE}} \approx 5 \times 10^{49}$ erg, a relativistic bulk $\Gamma \approx 100$, and expands into a low-density medium having a number density $n \approx 10^{-2} \text{ cm}^{-3}$ (see Fig. 2).

The accurate location of GRB 050509b by the BAT and XRT on board Swift led for the first time to the identification of the X-ray afterglow of a short GRB²⁷. However, the burst occurred in the direction of two merging clusters of galaxies, making it impossible to determine the host galaxy of the burst and thus the redshift of the burst (>20 galaxies lie within the XRT error circle for the X-ray

afterglow of the burst), let alone the location of the burst within the host galaxy.

The accurate location of GRB 050709 by the WXM and SXC on board HETE has led to the identification of the X-ray afterglow¹⁰, and for the first time, the identification in ground-based¹¹ and HST¹⁰ images of the optical afterglow of a short-hard burst. These have led to the first secure identification of a host galaxy: a late-type spiral galaxy lying at a redshift of $z = 0.16$ (ref. 11). The X-ray and optical afterglows lie at a projected distance of ~ 3 kpc from the centre of the host galaxy and are therefore not coincident with the brightest optical emission from the host galaxy (in contrast to long bursts²⁸).

These results constrain the nature of the central engine for GRB 050709 and, by implication, all short-hard bursts. The absence of any large-amplitude oscillations with a period in the range 1–10 s in the long-soft bump and the offset of GRB 050709 from the centre of the host galaxy argue against an association between this short-hard burst and SGR giant flares^{25,26}. Models based upon core collapse in massive stars⁹ explain the association of some long-soft bursts with supernovae. However, given the time it takes for the GRB-producing jet to emerge from a collapsing massive star, it is difficult for such models to produce bursts shorter than a few seconds. Merging neutron stars, on the other hand, can produce very short bursts²⁹. Given the smaller mass of the accretion disk that forms, it is not unreasonable to expect a lower average E_{iso} for short-hard bursts, and hence a smaller average redshift. Moreover, since binary neutron stars begin with a ‘kick’ at birth and travel large distances before merging, we expect an offset of the order of several kiloparsecs between the star-forming regions of the host galaxy and the burst³⁰. GRB 050709 exhibits all of these properties. The roughly 100 s lag between the short-hard pulse of GRB 050709 and the peak of the much softer afterglow is consistent with the low-density interstellar medium expected in the vicinity of a merging compact binary²². If short-hard GRBs are due to merging neutron stars, they produce powerful bursts of gravitational radiation that should be detectable by the second-generation Laser Interferometry Gravitational-wave Observatory.

The HETE localization of the short-hard burst GRB 050709 has led to follow-up observations that have identified the X-ray afterglow and (for the first time) the optical afterglow of a short, hard burst. These, in turn, have led to identification of the host galaxy of the burst as a late-type galaxy at $z = 0.16$, showing that at least some

Table 1 | Spectral model parameters for GRB 050709

Parameter	$t = 0-0.20$ s	$t = 0.20-0.50$ s	$t = 0-0.50$ s	$t = 20-180$ s
Spectral model	PLE	PLE	PLE	Power-law
Photon index, α	$0.53^{+0.12}_{-0.13}$	$0.55^{+1.0}_{-1.3}$	$0.82^{+0.13}_{-0.14}$	$1.98^{+0.18}_{-0.15}$
Peak energy ($E_{\text{peak}}^{\text{obs}}$)	$83.9^{+11}_{-8.3}$	$10.6^{+4.5}_{-3.5}$	86.5^{+16}_{-11}	—
Normalization (at 15 keV)	$0.79^{+0.07}_{-0.08}$	$0.650^{+3.24}_{-0.48}$	$0.377^{+0.04}_{-0.04}$	$0.0075^{+0.0013}_{-0.0013}$
χ^2 (d.o.f.)	439 (366)	374 (366)	467 (366)	336 (367)

The spectral models are power-law times exponential (PLE) and power-law. Errors are for 90% confidence. The normalization units are photons $\text{cm}^{-2} \text{s}^{-1} \text{keV}^{-1}$. d.o.f. is the number of degrees of freedom in the fit. We have also fitted power-law and Band models to the first three time intervals and find χ^2 (d.o.f.) values of 550 (367), 383 (367), and 538 (367) for the power-law model and 439 (365), 374 (365), and 467 (365) for the Band model, demonstrating that the data require the PLE model but not the Band model. Similarly, we have also fitted a PLE model to the fourth time interval and find χ^2 (d.o.f.) values of 336 (366), demonstrating that the data require the power-law model but not the PLE model. The large values of χ^2 per degree of freedom that we find for the PLE model for the 0–0.2 s and 0–0.5 s time intervals are due to the rapid spectral evolution during the short-hard pulse. We have calculated the hardness ratio $H = \text{counts (5–10 keV)}/\text{counts (2–5 keV)}$ for 10, 20, 30 and 40 s time intervals and find no evidence of spectral evolution during the long-soft bump. The time history and spectral parameters of GRB 050709 are similar to those of the BATSE bursts GRBs 921022, 990516 and 990712 (refs 17, 18). In particular, the short-hard pulse of GRB 921022 had a duration ~ 256 ms and spectral parameters $\alpha = -1.1 \pm 0.3$, $\beta = -2.15 \pm 0.12$, and $E_{\text{peak}}^{\text{obs}} = 123 \pm 28$ keV, and the long-soft bump of that burst had a power-law spectrum with index $\alpha \approx -2$ (ref. 17), while the short-hard pulse of GRB 990712 had a duration ~ 0.75 s and spectral parameters $\alpha \approx -0.2$ and $E_0 \approx 600$ keV, and the long-soft bump of that burst had a power-law spectrum with index $\alpha = -1.9 \pm 0.6$ (ref. 23).

Table 2 | Emission properties of GRB 050709

Energy (keV)	Peak photon flux (photons $\text{cm}^{-2} \text{s}^{-1}$)	Photon fluence (photons cm^{-2})	Peak energy flux ($10^{-8} \text{ erg cm}^{-2} \text{s}^{-1}$)	Energy fluence ($10^{-8} \text{ erg cm}^{-2}$)
Short pulse				
2–10	29.0 ± 5.2	3.47 ± 0.59	24.9 ± 3.9	2.81 ± 0.42
2–25	53.6 ± 6.1	5.55 ± 0.69	88.7 ± 7.7	8.31 ± 0.70
2–30	58.1 ± 6.9	5.94 ± 0.70	111 ± 8.6	10.1 ± 0.76
7–30	37.1 ± 2.8	3.31 ± 0.25	96.8 ± 6.8	8.35 ± 0.59
30–400	34.1 ± 2.7	2.51 ± 0.22	400 ± 46	30.3 ± 3.8
50–100	13.9 ± 1.1	0.986 ± 0.087	156 ± 13	11.0 ± 1.0
100–300	6.62 ± 1.1	0.515 ± 0.092	155 ± 29	12.4 ± 2.5
2–400	92.1 ± 7.6	8.43 ± 0.752	511 ± 49	40.3 ± 4.1
Long bump				
2–10	2.36 ± 0.43	107 ± 18	1.53 ± 0.27	69.1 ± 10
2–25	2.72 ± 0.47	123 ± 18	2.41 ± 0.37	109 ± 14

The quantities in this table are derived assuming the best-fit PLE model for the spectrum of the short-hard pulse and the best-fit PL model for the spectrum of the long-soft bump. Errors are for 90% confidence. The photon number and photon energy peak fluxes for the short-hard pulse were evaluated over a 70 ms interval, corresponding to T_{90} for the short-hard pulse; those for the long-soft bump are evaluated in a 1 s interval. T_{90} is the time interval containing 90% of the photons. Using the redshift $z = 0.16$ measured for the host galaxy¹¹, the isotropic-equivalent energy of the short-hard pulse in the 1–10,000 keV energy band in the rest frame of the source is $E_{\text{iso}} = (2.8^{+0.4}_{-0.2}) \times 10^{49}$ erg, taking $\Omega_M = 0.3$, $\Omega_\Lambda = 0.7$, and $h = 0.65$. Using a time interval 0.060 s in the rest frame of the source (corresponding to a duration of 0.07 s in the observer frame) and assuming the same cosmology and energy band, the luminosity of the short-hard pulse in the 1–10,000 keV energy band in the rest frame of the source is $L_{\text{iso}} = (5.2 \pm 0.7) \times 10^{50} \text{ erg s}^{-1}$.

short-hard bursts occur at cosmological distances in the outskirts of galaxies, and are likely to be caused by the merging of compact binaries.

Received 26 July; accepted 9 September 2005.

- Hurley, K. in *Gamma-Ray Bursts* (eds Paciesas, W. & Fishman, G.) 3 (AIP, New York, 1992).
- Lamb, D. Q., Graziani, C. & Smith, I. Evidence for two distinct morphological classes of gamma-ray bursts from their short time scale variability. *Astrophys. J.* **413**, L11–L14 (1993).
- Kouveliotou, C. et al. Identification of two classes of gamma-ray bursts. *Astrophys. J.* **413**, L101–L104 (1993).
- Costa, E. et al. Discovery of an X-ray afterglow associated with the gamma-ray burst of 28 February 1997. *Nature* **387**, 783–784 (1997).
- van Paradijs, J. et al. Transient optical emission from the error box of the gamma-ray burst of 28 February 1997. *Nature* **386**, 686–688 (1997).
- Metzger, M. R. et al. Spectral constraints on the redshift of the optical counterpart to the gamma-ray burst of 8 May 1997. *Nature* **387**, 878–879 (1997).
- Castander, F. & Lamb, D. Q. A photometric investigation of the GRB 970228 afterglow and the associated nebulosity. *Astrophys. J.* **523**, 593–601 (1999).
- Woosley, S. E. Gamma-ray bursts from stellar mass accretion disks around black holes. *Astrophys. J.* **405**, 273–277 (1993).
- Stanek, C. et al. Spectroscopic discovery of the supernova 2003dh associated with GRB 030329. *Astrophys. J.* **591**, L17–L20 (2003).
- Fox, D. B. et al. The afterglow of GRB 050709 and the nature of the short-hard γ -ray bursts. *Nature* doi:10.1038/nature04189 (this issue).
- Hjorth, J. et al. The optical afterglow of the short γ -ray burst GRB 050709. *Nature* doi:10.1038/nature04174 (this issue).
- Ricker, G. R. et al. in *Gamma-Ray Burst and Afterglow Astronomy 2001: A Workshop Celebrating the First Year of the HETE Mission* (eds Ricker, G. R. & Vanderspek, R. K.) 3–16 (AIP Press, New York, 2003).
- Paciesas, W. et al. in *Gamma-Ray Bursts in the Afterglow Era* (eds Costa, E., Frontera, F. & Hjorth, J.) 13–15 (Springer, Berlin, 2000).
- Ghirlanda, G., Ghisellini, G. & Celotti, A. The spectra of short gamma-ray bursts. *Astron. Astrophys.* **422**, L55–L58 (2004).
- Lamb, D. Q. et al. in *Gamma-Ray Bursts in the Afterglow Era* (eds Feroci, M., Frontera, F., Masetti, N. & Piro, L.) 94–97 (ASP, San Francisco, 2004).
- Paciesas, W. et al. The Fourth BATSE Gamma-Ray Burst Catalog (revised). *Astrophys. J. Suppl.* **122**, 465–495 (2000).
- Preece, R. et al. The BATSE Gamma-Ray Burst Spectral Catalog. I. High time resolution spectroscopy of bright bursts using high energy resolution data. *Astrophys. J. Suppl.* **126**, 19–36 (2000).
- Frail, D. A. et al. Beaming in gamma-ray bursts: Evidence for a standard energy reservoir. *Astrophys. J.* **562**, L55–L58 (2001).
- Amati, L. et al. Intrinsic spectra and energetics of BeppoSAX gamma-ray bursts with known redshifts. *Astron. Astrophys.* **390**, 81–89 (2002).
- Lamb, D. Q. et al. Scientific highlights of the HETE-2 mission. *New Astron. Rev.* **48**, 423–430 (2004).
- Yonetoku, D. et al. Gamma-ray burst formation rate inferred from the spectral peak energy-peak luminosity relation. *Astrophys. J.* **609**, 935–951 (2004).
- Lazzati, D., Ramirez-Ruiz, E. & Ghisellini, G. Possible detection of hard X-ray afterglows of short γ -ray bursts. *Astron. Astrophys.* **379**, L39–L43 (2001).
- Frederiks, D. D. et al. in *Gamma-Ray Bursts in the Afterglow Era* (eds Feroci, M., Frontera, F., Masetti, N. & Piro, L.) 197–200 (ASP, San Francisco, 2004).
- Connaughton, V. BATSE observations of gamma-ray burst tails. *Astrophys. J.* **567**, 1028–1036 (2002).
- Hurley, K. et al. An exceptionally bright flare from SGR 1806-20 and the origin of short duration gamma-ray bursts. *Nature* **434**, 1098–1103 (2005).
- Palmer, D. M. et al. A giant gamma-ray flare from the magnetar SGR 1806-20. *Nature* **434**, 1107–1109 (2005).
- Gehrels, N. et al. A short γ -ray burst apparently associated with an elliptical galaxy at redshift $z = 0.225$. *Nature* doi:10.1038/nature04142 (this issue).
- Bloom, J. et al. The observed offset distribution of gamma-ray bursts from their host galaxies: A robust clue to the nature of the progenitors. *Astron. J.* **123**, 1111–1148 (2002).
- Eicher, D., Livio, M., Piran, T. & Schramm, D. N. Nucleosynthesis, neutrino bursts and γ -rays from coalescing neutron stars. *Nature* **340**, 126–128 (1989).
- Fryer, C. L., Woosley, S. E. & Hartmann, D. H. Formation rates of black hole accretion disk gamma-ray bursts. *Astrophys. J.* **526**, 152–177 (1999).
- Reichart, D. E. et al. A possible Cepheid-like luminosity estimator for the long gamma-ray bursts. *Astrophys. J.* **552**, 57 (2001).

Acknowledgements This research was supported in the USA by NASA.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to G.R.R. (grr@space.mit.edu) or J.S.V. (jsvilla@space.mit.edu).

The optical afterglow of the short γ -ray burst GRB 050709

Jens Hjorth¹, Darach Watson¹, Johan P. U. Fynbo¹, Paul A. Price², Brian L. Jensen¹, Uffe G. Jørgensen¹, Daniel Kubas³, Javier Gorosabel⁴, Páll Jakobsson¹, Jesper Sollerman^{1,5}, Kristian Pedersen¹ & Chryssa Kouveliotou⁶

It has long been known that there are two classes¹ of γ -ray bursts (GRBs), mainly distinguished by their durations. The breakthrough in our understanding of long-duration GRBs (those lasting more than ~ 2 s), which ultimately linked them with energetic type Ic supernovae^{2–4}, came from the discovery of their long-lived X-ray⁵ and optical^{6,7} ‘afterglows’, when precise and rapid localizations of the sources could finally be obtained. X-ray localizations have recently become available^{8,9} for short (duration < 2 s) GRBs, which have evaded optical detection for more than 30 years. Here we report the first discovery of transient optical emission (R-band magnitude ~ 23) associated with a short burst: GRB 050709. The optical afterglow was localized with subarcsecond accuracy, and lies in the outskirts of a blue dwarf galaxy. The optical and X-ray¹⁰ afterglow properties 34 h after the GRB are reminiscent of the afterglows of long GRBs, which are attributable to synchrotron emission from ultrarelativistic ejecta. We did not, however, detect a supernova, as found in most nearby long GRB afterglows, which suggests a different origin for the short GRBs.

NASA’s High Energy Transient Explorer, HETE-2, detected GRB 050709 at 22:36:37 UT on 9 July 2005 (ref. 9). The burst consisted of a single hard γ -ray pulse only 70 ms long, with peak energy of 83 keV, followed ~ 30 s later by fainter and softer emission from the same location, detected only at energies below 10 keV and lasting for about 2 min. This later component, speculated to be due to afterglow emission⁹, allowed HETE-2’s Soft X-ray Camera (SXC) to obtain a source location with an 81-arcsec error radius.

The properties of GRB 050709 place it firmly in the elusive short GRB population, as observed¹ by the Burst And Transient Source Experiment (BATSE). The first, hard pulse was shorter than most BATSE short-hard GRBs and had a peak energy and spectral slope consistent with the short-hard distribution^{1,9,11} (Fig. 1). However, the second pulse of faint, soft X-ray emission, observed only by SXC, would not have been detectable with the BATSE Large Area Detectors, which had a low energy threshold of ~ 30 keV. Thus the entire event would have been classified as a short GRB in the BATSE sample. Moreover, we note that statistical studies that summed the background-subtracted signals of large numbers of BATSE bursts have also shown long-lasting (hundreds of seconds) fainter emission following the main event^{12,13}.

We observed the SXC error circle of GRB 050709 with the Danish 1.5-m telescope at the La Silla Observatory starting 33 h after the event, and obtained images in good observing conditions over the following 18 days (Table 1). We used point-spread-function-matched image subtraction¹⁴ and established the existence of a single transient

source inside the error circle positionally coincident with an X-ray source¹⁰ observed with the Chandra X-ray Observatory (CXO) (Fig. 2). The point source is positioned at right ascension $\alpha_{2000} = 23$ h 01 min 26.957 s, declination $\delta_{2000} = -38^\circ 58' 39.76''$, with an error of 0.25 arcsec, and sits on the edge of an extended source, apparently the host galaxy of the burst. Moreover, the source faded with a decay index ($f_\nu \propto t^\alpha$, where f_ν is the flux density and t is time) of $\alpha = -1.33 \pm 0.45$ (Fig. 3), similar to the behaviour of afterglow emission from long GRBs. We therefore conclude that this source is the optical counterpart to the short GRB 050709. The

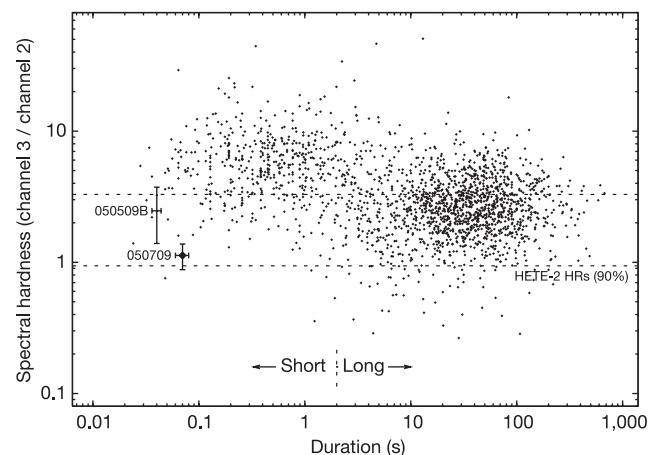


Figure 1 | The classic BATSE duration-spectral hardness diagram¹.

GRB 050709 (large filled diamond) is clearly among the shortest GRBs, more than an order of magnitude below the canonical 2 s divide between long and short bursts. The burst is also among the softer short GRBs. This fact is probably an instrumental bias effect related to the soft response function of HETE-2’s WXM, which is sensitive in the 2–25 keV bandpass (compared to BATSE’s nominal trigger range between 50 and 300 keV). Indeed, most long GRBs localized by WXM have soft spectra—the 90% limits for the hardness ratios (HRs) of a sample of 50 long FREGATE/WXM detected GRBs are plotted as dashed lines. The HETE-2 hardness ratios were converted to the BATSE spectral 50–100 keV (channel 2) and 100–300 keV (channel 3) bands assuming a simple power-law model. The only short GRB rapidly localized so far by the Swift-XRT, GRB 050509B, is plotted for comparison. The hardness ratio for GRB 050709 was derived from an exponentially cut-off power-law model fit to the data from the short, hard pulse⁹. (The long, soft component that could not have been detected by BATSE is not considered.) Error bars are 68% confidence limits.

¹Dark Cosmology Centre, Niels Bohr Institute, University of Copenhagen, Juliane Maries Vej, DK-2100 Copenhagen Ø, Denmark. ²Institute for Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, Hawaii 96822, USA. ³ESO Santiago, Casilla 19001, Santiago 19, Chile. ⁴Instituto de Astrofísica de Andalucía (IAA-CSIC), PO Box 3004, E-18080 Granada, Spain. ⁵Department of Astronomy, Stockholm University, AlbaNova, 106 91 Stockholm, Sweden. ⁶NASA/Marshall Space Flight Center, National Space Science Technology Center, XD-12, 320 Sparkman Drive, Huntsville, Alabama 35805, USA.

Table 1 | Danish 1.5-m optical imaging observations of GRB 050709

Date (July 2005 UT)	Band	Int. time (s)	Seeing (FWHM in arcsec)	Afterglow flux (μ Jy)
11.3587	R	12 × 600	0.7	2.34 ± 0.12
12.3283	R	17 × 600	1.0	1.17 ± 0.26
17.3233	R	24 × 600	1.1	−0.06 ± 0.23
18.3999	R	13 × 600	0.9	0
19.3991	V	3 × 600	1.6	
19.4062	I	2 × 300 + 2 × 600	1.4	
19.4106	B	3 × 600	1.6	
27.3799	R	12 × 600	1.0	0.21 ± 0.17
29.3957	R	8 × 600	1.5	0.13 ± 0.24

GRB 050709 occurred on July 9.9421 2005 UT. The images were bias-subtracted and flat-fielded in the standard manner. Aperture photometry of the source is complicated by the presence of the bright ($R \approx 21.2$ mag) galaxy, so we used image subtraction¹⁴ to properly measure the afterglow fluxes. The reported fluxes are relative to the R-band image obtained on July 18.4 UT. Extrapolation of the power-law fit to the first data points suggests an afterglow flux at this time of 0.16 μ Jy, consistent, within errors, with zero flux. In this case the relative fluxes become absolute fluxes as assumed in Fig. 3.

extended source is an emission-line galaxy at a redshift $z = 0.16$ (ref. 10). Assuming a flat Universe with Hubble constant $H_0 = 72 \text{ km s}^{-1} \text{ Mpc}^{-1}$, matter density parameter $\Omega_m = 0.27$ and equation of state parameter $w = -1$ gives a luminosity distance $d_L = 747 \text{ Mpc}$ and a projected distance of the transient from the galaxy centre of $\sim 3.5 \text{ kpc}$ ($1.3 \pm 0.1 \text{ arcsec}$), strongly suggesting that this is the host galaxy of the GRB. At this distance, the fluence of the short pulse converts to an isotropic-equivalent energy release of $2.2 \times 10^{49} \text{ erg}$ (ref. 9). Additional multicolour imaging obtained at the Danish 1.5-m telescope (Table 1) shows that the host is a fairly blue dwarf galaxy with B-, V-, R- and I-band magnitudes of 22.22 ± 0.10 , 21.22 ± 0.10 , 21.24 ± 0.12 and 20.38 ± 0.25 , respectively (uncorrected for the small Galactic extinction). At $z = 0.16$ these colours indicate that the host is a starburst galaxy with little extinction, an age of 0.4 Gyr, a total luminosity of approximately $0.03L^*$ (where L^* is the characteristic luminosity of the Schechter luminosity function) (blue absolute magnitude $M_B = -16.9 \pm 0.1$), and a star formation rate of the order of $0.1 M_\odot \text{ yr}^{-1}$ (where $M_\odot$ is the solar mass).

Our observations thus establish that the short-duration GRB 050709 was accompanied by long-lived optical afterglow emission, and most probably occurred in the outskirts of a star-forming dwarf galaxy. In these respects, GRB 050709 shares the properties of long-duration GRBs. However, several other properties, in addition to its short duration, set this event apart from long GRBs. The prompt emission⁹ is less luminous than most long GRBs. In particular, GRB 050709 does not lie close to the luminosity–peak energy correlation established for long bursts¹⁵—its peak energy ($\sim 83 \text{ keV}$)⁹ is much higher than observed in the few long GRBs that are this faint. Likewise, the X-ray afterglow¹⁰ was about 1,000 times fainter than the equivalent isotropic X-ray luminosities of long GRB afterglows¹⁶.

What do our optical observations tell us about the progenitor of GRB 050709? Short GRB progenitor models¹⁷ include the merger of two compact objects, such as two neutron stars¹⁸ or a neutron star and a black hole, or, in the ‘unification scheme’^{19,20}, a variant of the collapsar model.

Our upper limits on optical emission between 7 and 20 days after the burst (Table 1) rule out the occurrence of an energetic type Ic supernova (SN) as found in most long GRBs at low redshifts. To accommodate the non-detection in our data (which corresponds to an absolute magnitude of $M_V > -14.7$ at 16 days in the restframe), any supernova must have been at least 7 times fainter than the faint type Ic SN 1994I (Fig. 3). A supernova similar to the canonical GRB 980425/SN 1998bw² would have reached a brightness of $R = 19.6 \text{ mag}$, as was also observed in the long GRB 030329/SN 2003dh at $z = 0.17$ (ref. 3), that is, 4.6 mag or 70 times brighter than the derived upper limits. Remarkably, both GRB 050509B, the only short GRB that had previously been localized to better than

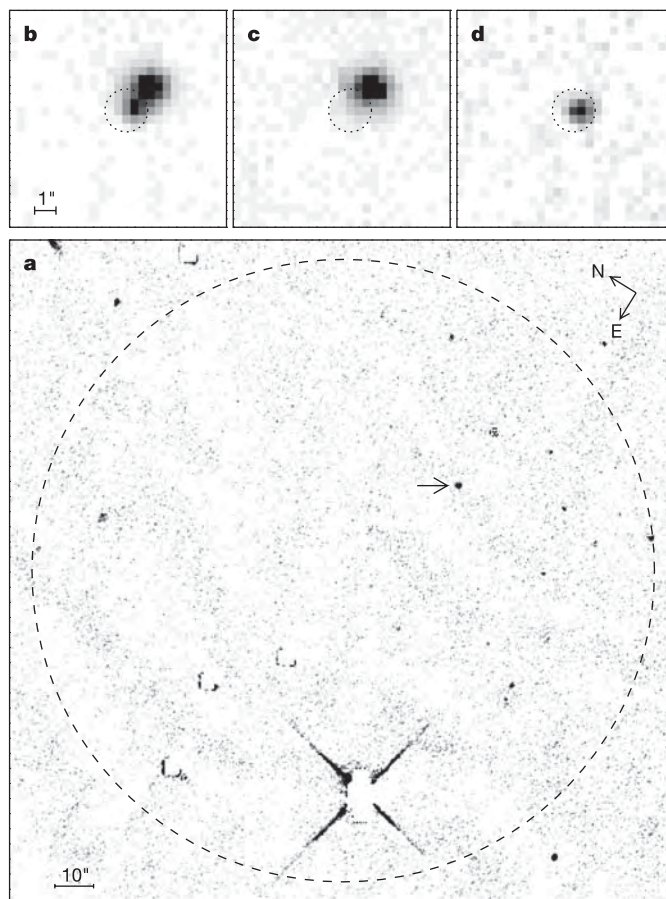


Figure 2 | The optical afterglow of GRB 050709. **a**, Variable sources in an R-band image of the field of GRB 050709 taken with the Danish 1.5-m telescope + DFOSC at La Silla on 11 July 2005, commencing at 6:29 UT. Subtraction of a reference frame taken a week later reveals that only one source inside the indicated HETE-2 SXC error circle⁹, identified as the afterglow of GRB 050709, faded away. The subtraction was performed by solving for the best convolution kernel between stars common to the two images¹⁴ using discrete basis functions to describe the kernel. The optical source (arrowed) is situated on the edge of a low-redshift galaxy. **b**, The optical R-band image from July 11.4 UT and the CXO 95% error circle. **c**, The July 18.4 UT R-band image showing the host galaxy after the afterglow has faded away. **d**, The difference between the July 11.4 UT and the July 18.4 UT images showing the afterglow that has faded away over this period. The orientation and scale of the images are indicated.

arcminute precision^{8,21}, and GRB 050709 were located in low-redshift galaxies^{10,21}, that is, suitable for a supernova search, but in neither case was evidence for a supernova found in their light curves²². Although we cannot entirely rule out the possibility that GRB 050509B originated from a high-redshift star-forming galaxy beyond its presumed host galaxy, this is very unlikely for GRB 050709. Bright SN 1998bw-like core-collapse supernovae are, therefore, effectively ruled out as the progenitors of short GRBs. In the unification (collapsar) scheme, a supernova similar to those seen in long GRBs is predicted¹⁹. Unless a mechanism to suppress their supernova signals is found, our results disfavour⁵⁶ Ni-producing collapsars.

The optical afterglow was located in the outskirts of a star-forming, subluminescent host galaxy. Compact object mergers are expected to occur in a more uniformly distributed fashion about their host galaxies than are collapsars, as a result of the evolutionary merging timescale involved (Myr–Gyr)²³. The compact binary may travel several kiloparsecs before the merger owing to the kick imparted to the binary during the supernova explosions of the progenitor stars. Collapsars occur in starburst galaxies²⁴, whereas

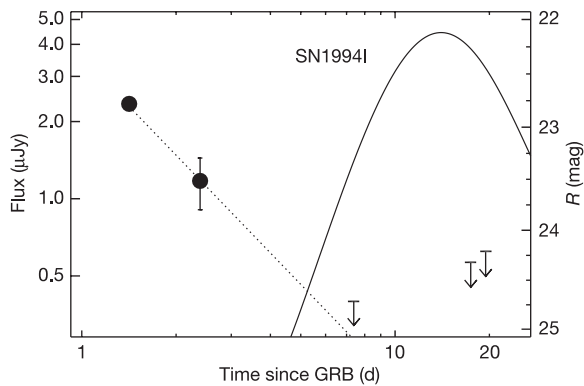


Figure 3 | Light curve of the optical counterpart to GRB 050709. The data points are from Table 1. Error bars are 1σ standard deviations. Upper limits are 2σ upper limits. Fitting a power law to the first two data points (dotted line) yields a decay slope of $\alpha = -1.33 \pm 0.45$ (including the first upper limit results in $\alpha = -1.61 \pm 0.44$). This value is similar to the decay slopes observed for long GRBs, and is indicative of synchrotron emission. Also shown (solid curve) is the light curve of the faint type Ic SN 1994I, which is fainter than normal template supernovae of types Ia and Ic²². Our upper limits at 18 and 20 days clearly rule out even such a faint supernova.

mergers of two neutron stars are expected to have a wide distribution in coalescence timescales²³ and so are expected to occur in both old and young galaxies. We note that the offset of the burst position from its host galaxy centre is among the largest observed²⁵, and the afterglow is not conspicuously centred on a bright part of the host as is usually seen in long GRBs²⁶. Whereas GRB 050509B was associated with an elliptical galaxy, our observations of a starburst host galaxy establishes the diverse nature of these events. This is analogous to the way type Ia supernovae are found in diverse environments. The different environments and the different mean ages of the stellar population may be indicative of a wide range of formation timescales (for example, the in-spiral timescale in the merger model) or simply reflect the likely existence of an old population in starburst galaxies.

At 34 h, the afterglow exhibits a very steep optical to X-ray spectral index, $\beta_{\text{OX}} = -1.1 \pm 0.1$ ($f_{\nu} \propto \nu^{\beta}$), steeper than that observed for any afterglow of a long GRB²⁷. A merger of two neutron stars may lead to a 'mini-supernova'—that is, emission in the ultraviolet/optical domain at about 1 d after the merger with peak brightness similar to a supernova, and arising from subrelativistic ejecta condensing to form neutron-rich, radioactive material that acts as a heat source for an expanding envelope²⁸. In view of the steep optical to X-ray spectral index, our detection of an optical transient may be consistent with a variant of the mini-supernova model; we note, however, that this probably requires the X-ray and optical emission to originate from different emitting regions. Alternatively, the afterglow can be attributed to synchrotron emission from relativistic ejecta¹⁷, much as in long GRBs^{12,29} and for the X-ray afterglow of GRB 050509B^{8,21,30}. The observed parameters for GRB 050709, $\alpha = -1.33 \pm 0.45$ and $\beta_{\text{OX}} = -1.1 \pm 0.1$, are consistent with synchrotron emission with a power-law electron energy distribution with index $2.0 < p < 3.4$. Synchrotron emission from short GRBs is expected to be much fainter than long GRB afterglows as the total energy deposited is expected to be smaller. The much fainter X-ray afterglow of GRB 050509B and the lack of optical detection may be related to its smaller isotropic energy output ($\sim 1.1 \times 10^{48}$ erg; ref. 8), ~ 20 times less than in GRB 050709.

The localization of short GRBs have lead to intriguing observations that seem to support the leading merger model^{8,21,30}. However, with a sample of only two events it would be prudent not to draw definitive conclusions at this stage about the progenitors of short GRBs. Future optical observations, even from small telescopes,

will play an important role in finally determining the origin of short GRBs, although ultimately, detection of gravitational waves²³ will demonstrate whether short GRBs result from mergers of coalescing binary compact stars.

Received 24 July; accepted 26 August 2005.

- Kouveliotou, C. *et al.* Identification of two classes of gamma-ray bursts. *Astrophys. J.* **413**, L101–L104 (1993).
- Galama, T. J. *et al.* An unusual supernova in the error box of the γ -ray burst of 25 April 1998. *Nature* **395**, 670–672 (1998).
- Hjorth, J. *et al.* A very energetic supernova associated with the γ -ray burst of 29 March 2003. *Nature* **423**, 847–850 (2003).
- Stanek, K. Z. *et al.* Spectroscopic discovery of the supernova 2003dh associated with GRB 030329. *Astrophys. J.* **591**, L17–L20 (2003).
- Costa, E. *et al.* Discovery of an X-ray afterglow associated with the γ -ray burst of 28 February 1997. *Nature* **387**, 783–785 (1997).
- Van Paradijs, J. *et al.* Transient optical emission from the error box of the γ -ray burst of 28 February 1997. *Nature* **368**, 686–688 (1997).
- Metzger, M. R. *et al.* Spectral constraints on the redshift of the optical counterpart to the γ -ray burst of 8 May 1997. *Nature* **387**, 878–879 (1997).
- Gehrels, N. *et al.* A short γ -ray burst apparently associated with an elliptical galaxy. *Nature* doi:10.1038/nature04142 (this issue).
- Villasenor, J. S. *et al.* Discovery of the short γ -ray burst GRB 050709. *Nature* doi:10.1038/nature04213 (this issue).
- Fox, D. B. *et al.* The afterglow of GRB 050709 and the nature of the short-hard γ -ray bursts. *Nature* doi:10.1038/nature04189 (this issue).
- Ghirlanda, G., Ghisellini, G. & Celotti, A. The spectra of short gamma-ray bursts. *Astron. Astrophys.* **422**, L55–L58 (2004).
- Lazzati, D., Ramirez-Ruiz, E. & Ghisellini, G. Possible detection of hard X-ray afterglows of short gamma-ray bursts. *Astron. Astrophys.* **379**, L39–L43 (2001).
- Connaughton, V. BATSE observations of gamma-ray burst tails. *Astrophys. J.* **567**, 1028–1036 (2002).
- Barris, B. J., Tonry, J. L., Novicki, M. C. & Wood-Vasey, W. M. The NN2 flux difference method for constructing variable object light curves. *Astron. J.* (in the press); preprint at (<http://arXiv.org/astro-ph/0507584>) (2005).
- Amati, L. *et al.* Intrinsic spectra and energetics of BeppoSAX gamma-ray bursts with known redshifts. *Astron. Astrophys.* **390**, 81–89 (2002).
- Berger, E. *et al.* A standard kinetic energy reservoir in gamma-ray burst afterglows. *Astrophys. J.* **590**, 379–385 (2003).
- Piran, T. The physics of gamma-ray bursts. *Rev. Mod. Phys.* **76**, 1143–1210 (2005).
- Rossow, S. From neutron star binaries to gamma-ray bursts. Preprint at (<http://arXiv.org/astro-ph/0504368>) (2005).
- Zhang, W., Woosley, S. E. & MacFadyen, A. I. Relativistic jets in collapsars. *Astrophys. J.* **586**, 356–371 (2003).
- Yamazaki, R., Ioka, K. & Nakamura, T. A unified model of short and long gamma-ray bursts, X-ray-rich gamma-ray bursts, and X-ray flashes. *Astrophys. J.* **607**, L103–L106 (2004).
- Bloom, J. S. *et al.* Closing in on a short-hard burst progenitor: constraints from early-time optical imaging and spectroscopy of a possible host galaxy of GRB 050509b. *Astrophys. J.* (in the press); preprint at (<http://arXiv.org/astro-ph/0505480>) (2005).
- Hjorth, J. *et al.* GRB 050509B: Constraints on short gamma-ray burst models. *Astrophys. J.* **630**, L117–L120 (2005).
- Voss, R. & Tauris, T. M. Galactic distribution of merging neutron stars and black holes—prospects for short gamma-ray burst progenitors and LIGO/VIRGO. *Mon. Not. R. Astron. Soc.* **342**, 1169–1184 (2003).
- Christensen, L., Hjorth, J. & Gorosabel, J. UV star-formation rates of GRB host galaxies. *Astron. Astrophys.* **425**, 913–926 (2004).
- Bloom, J. S., Kulkarni, S. R. & Djorgovski, S. G. The observed offset distribution of gamma-ray bursts from their host galaxies: A robust clue to the nature of the progenitors. *Astron. J.* **123**, 1111–1148 (2002).
- Fruchter, A. S. *et al.* The locations of cosmic explosions. *Nature* (submitted).
- Jakobsson, P. *et al.* Swift identification of dark gamma-ray bursts. *Astrophys. J.* **617**, L21–L24 (2004).
- Li, L. X. & Paczyński, B. Transient events from neutron star mergers. *Astrophys. J.* **507**, L59–L62 (1998).
- Panaitescu, A., Kumar, P. & Narayan, R. Observational prospects for afterglows of short-duration gamma-ray bursts. *Astrophys. J.* **561**, L171–L174 (2001).
- Lee, W. H., Ramirez-Ruiz, E. & Granot, J. A compact binary merger model for the short, hard GRB 050509b. *Astrophys. J.* **630**, L165–L168 (2005).

Acknowledgements We thank T. Tauris for discussions. The Dark Cosmology Centre is funded by the DNRF. We acknowledge benefits from collaboration within the EU FP5 Research Training Network 'Gamma-ray bursts: an enigma and a tool'.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.H. (jens@astro.ku.dk).

LETTERS

Microscopic artificial swimmers

R mi Dreyfus¹, Jean Baudry¹, Marcus L. Roper², Marc Fermigier³, Howard A. Stone² & J r me Bibette¹

Microorganisms such as bacteria and many eukaryotic cells propel themselves with hair-like structures known as flagella, which can exhibit a variety of structures and movement patterns¹. For example, bacterial flagella are helically shaped² and driven at their bases by a reversible rotary engine³, which rotates the attached flagellum to give a motion similar to that of a corkscrew. In contrast, eukaryotic cells use flagella that resemble elastic rods⁴ and exhibit a beating motion: internally generated stresses give rise to a series of bends that propagate towards the tip^{5–7}. In contrast to this variety of swimming strategies encountered in nature, a controlled swimming motion of artificial micrometre-sized structures has not yet been realized. Here we show that a linear chain of colloidal magnetic particles linked by DNA and attached to a red blood cell can act as a flexible artificial flagellum. The filament aligns with an external uniform magnetic field and is readily actuated by oscillating a transverse field. We find that the actuation induces a beating pattern that propels the structure, and that the external fields can be adjusted to control the velocity and the direction of motion.

To achieve controlled motion or swimming of manmade microstructures, two conditions need to be fulfilled. First, energy should be injected and transferred into a mechanical deformation of the device. Second, the sequence of deformations must be cyclic and not time-reversible. The second requirement arises from the fact that fluid dynamics at the micrometre scale is dominated by viscous rather than inertial terms. This means that a purely reversible internal displacement is not associated with any net motion (the ‘scallop theorem’)⁸, and that a propelled microscopic device must thus use a swimming strategy that breaks the time-reversal invariance^{8–10}. In the case of spermatozoa, for example, a bending wave propagates from the head towards the tail, inducing a net translational velocity. The issue of the direction of this motion is related to the structure of the filament. More precisely, the filament surface governs the ratio of the tangential viscous coefficient $\zeta_{\parallel}$ to the perpendicular viscous coefficient $\zeta_{\perp}$ (refs 6, 7). For smooth flagella, slender-body theory¹¹ gives $\zeta_{\perp}/\zeta_{\parallel} \approx 2$, so that the wave and overall velocity have opposite directions. By contrast, for flagella decorated with mastigonemes (hair-like filaments), the effective tangential resistive coefficient becomes much larger than the effective perpendicular resistive coefficient¹², so that flagella carrying mastigonemes swim in the same direction as the propagating wave. The propulsion exerted by microorganisms has motivated some theoretical and numerical modelling^{6,9,13}, especially for natural flexible filaments^{14,15}. For this kind of swimmer, one relevant dimensionless parameter, previously called the ‘sperm number’¹⁶, is shown to be:

$$S_p = L / \left(\frac{\kappa}{\zeta_{\perp} \omega} \right)^{1/4}$$

where L is the length of the filament, κ the bending rigidity, and ω the angular frequency of the driving. S_p represents the relative importance of viscous to elastic stresses on the filament. For the so-called

‘one-armed swimmer’ two regimes were predicted: one dominated by internal elasticity at low S_p and the other dominated by viscous friction at high S_p . Between these two regimes, the existence of a maximum normalized swimming speed $V/L\omega$ was predicted for S_p of the order of unity. At this time, spermatozoa provide the only available experimental data and, using current experimental estimates of their stiffness¹⁶, seem to operate at $S_p = 7$. Motivated by these two regimes and their potential involvement in the propulsion mechanism of real flagella, we attempt to measure the swimming velocity of a controlled manmade swimmer as a function of the dimensionless number S_p and ask whether there is indeed a maximum and how it depends on the type of actuation.

Our microscopic device consists of two parts: a magnetically activated ‘flagellum’ that provides propulsion, and a second part, which is the object to be transported (a red blood cell is the example shown here). The flagellum is made of micrometre-scale magnetic colloids attached together by short flexible joints. The filament is driven by a time-varying magnetic field that transfers energy and leads to motion. Indeed, the filament tends to follow the field direction, and so by oscillating the orientation of the net magnetic field we actuate the filament. However, as a consequence of its flexibility and viscous friction, reorientation is also associated with bending. The coupling between magnetic forces, filament flexibility and the viscous drag from the solvent that acts on the filament generates controlled deformations. With the proper field conditions and flexibility, the time-reversal invariance of the filament deformations is broken, and thus the attached cell is propelled.

The filaments are made of superparamagnetic 1- μm -diameter colloids linked^{17,18} by several 107-nm-long (315 base pairs, bp) double-stranded DNAs as sketched in Fig. 1. The length and number of the DNA linkers and the particle diameter control the flexibility. These filaments are attached by one end to a red blood cell following the sample preparation procedure described in the ‘Sample preparation’ section of the Methods. To generate motion in a controlled direction and with a controlled speed, two magnetic fields are used. A

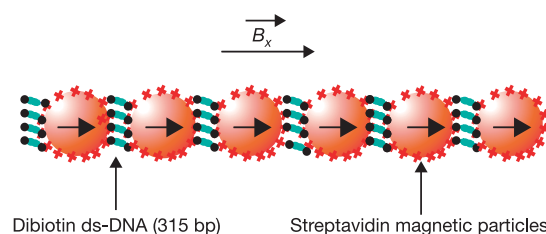


Figure 1 | Schematic representation of a flexible magnetic filament. The magnetic particles are coated with streptavidin (red cross symbols). Under an applied magnetic field B_x the particles form filaments. Double-stranded DNA with biotin at each end can bind the particles together via the specific biotin–streptavidin interaction. The experiments are performed with 8.4×10^4 DNAs per particle.

¹Laboratoire Collo ides et Mat riaux Divis s, ESPCI, UMR CNRS 7612 UPMC, ParisTech, 10 rue Vauquelin, 75005 Paris, France. ²Division of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA. ³Laboratoire Physique et M canique des Milieux H t rog nes, ESPCI, UMR CNRS 7636, ParisTech, 10 rue Vauquelin, 75005 Paris, France.

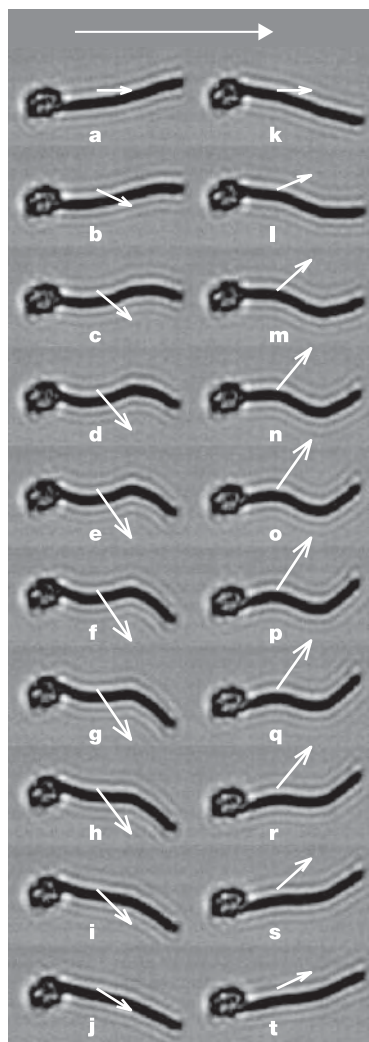


Figure 2 | Beating pattern of the motion of a magnetic flexible filament attached to a red blood cell. The time interval between each image is 5 ms. The white arrows represent the magnetic field ($B_x = 8.3$ mT, $B_y = 13.7$ mT, $f = 10$ Hz). The white arrow on the top shows the direction of motion. The filament length is $L = 24$ μ m.

homogeneous static field $\mathbf{B}_x = B_x \mathbf{x}$ imposes a straight configuration to the filament. In addition, a sinusoidal field $\mathbf{B}_y = B_y \sin(2\pi ft) \mathbf{y}$ with an adjustable frequency f is applied in the direction perpendicular to $\mathbf{B}_x$. These two fields have comparable amplitude so that the resulting field $\mathbf{B}_e$ oscillates around the x axis.

Because the particles are superparamagnetic, they acquire a magnetic dipole when subjected to a magnetic field. As the beads are known to have a preferred magnetization direction¹⁹, there are two different contributions to the magnetic torque exerted upon the filament: the first contribution is due to the dipolar interactions between the beads, and the second contribution is due to the interaction between the dipole and the external field. Both effects cause the filament to pivot to follow the magnetic field. We use a fast camera to record the dynamics of a microstructure formed by a red blood cell specifically bound to one end of a 30- μ m-long filament. As the transverse field $\mathbf{B}_y$ oscillates, the free end of the filament successively bends to follow the resulting field, which creates an undulation that propagates towards the attachment point.

In Fig. 2a–t we show a sequence of pictures of the dynamics during one period. The wave propagation, which is responsible for the non-reversible displacements of the filament, is visible in Fig. 3 in which we have superimposed the skeleton of each profile. We find that the

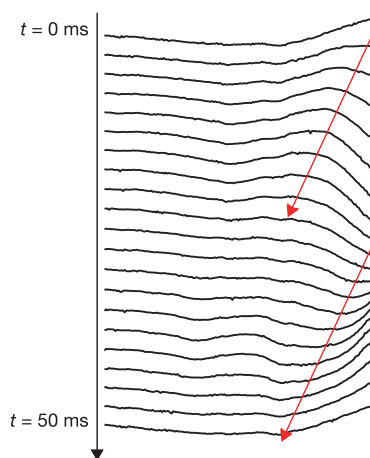


Figure 3 | Sequence of deformation of the end of a free filament. The propagation of a bending wave is indicated by the arrows. Conditions for the magnetic field: $B_x = 9$ mT, $B_y = 14$ mT, $f = 20$ Hz. Each image corresponds to pictures taken every 2.5 ms. The length of the portion of the filament shown is 34 μ m.

overall displacement of this swimmer is always directed towards the free extremity of the filament (see the Supplementary Movies) and opposite to the propagation of the bending wave⁷. This behaviour is an intrinsic consequence of these magnetically actuated flagella. The bending wave propagates from the free end because it is the most mobile and has the largest-amplitude displacement although the entire filament is magnetically actuated. By contrast, spermatozoa for which the bending wave propagates from head to tail move in the direction of the head. The ‘roughness’ of the filament evidently has no decisive impact upon the swimming direction, which, just as for smooth flagellum, opposes the direction of bending-wave propagation. This is in agreement with previous experiments and simulations^{20,21} that prove that a chain of spheres has almost the same drag coefficient as a prolate spheroid of the same length and aspect ratio. The mean swimming velocity of this device is always found collinear to the static field $\mathbf{B}_x$ and has a maximum of about the diameter of a red blood cell per second. As a final remark, we note that within the same sample we have sometimes observed two different swimming devices moving in opposite directions according to their initial orientation, which rules out any suspected consequences of residual field gradients. The Supplementary Movies show the dynamics of the filament at two frame rates: 440 frames s^{-1} (Supplementary Movie 1) and 40 frames s^{-1} (Supplementary Movie 2), for which $L = 12$ μ m, $f = 10$ Hz, $B_x = 9$ mT, $B_y = 14.5$ mT.

The physics of the motion of the magnetic filament can be described by the equations briefly presented in the ‘Equation of motion’ section of the Methods and detailed in the Supplementary Information. The final equations of motion involve three dimensionless numbers: S_p as previously defined, $b_0 = B_y/B_x$, and the magneto-elastic number:

$$M_n = \frac{2\pi(aB_xL)^2}{3\mu_0\kappa} \left(\frac{\chi_{||} - \chi_{\perp} + \chi_{||}\chi_{\perp}/4}{(1 - \chi_{||}/6)(1 + \chi_{\perp}/12)} \right)$$

where $\chi_{||}$ is the susceptibility of the easy direction, $\chi_{\perp}$ the susceptibility in the orthogonal direction, a the radius of the particles, and μ_0 the magnetic permeability in free space. To measure κ and estimate M_n , we assume the susceptibilities ($\chi_{||}$ and $\chi_{\perp}$) to be identical and equal to the value given by the manufacturer (see ‘Parameters measurement’ section in the Methods).

In Fig. 4, we plot the measured scaled swimming velocity $V/L\omega$ as a function of S_p , for three different values of M_n . As predicted^{14,15}, the normalized velocity has a maximum; however, it is clear that its

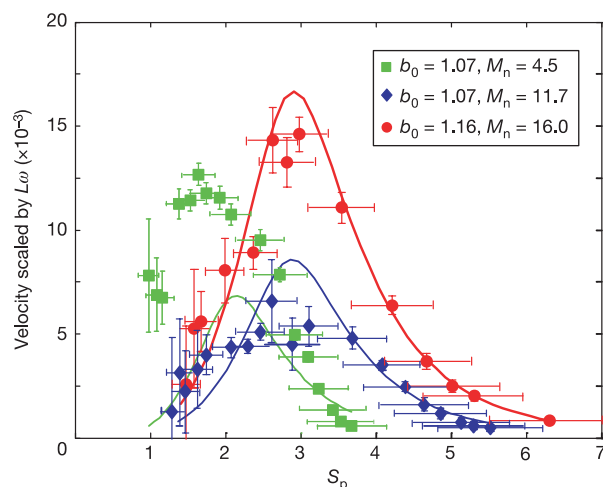


Figure 4 | Scaled velocity as a function of S_p . Experimental data are discrete points while the continuous lines are the predictions obtained with our model. Error bars for the scaled velocity are estimated by measuring the drift velocity of a filament when $B_y = 0$ and error bars for S_p correspond to the standard deviation of S_p calculated by estimating the standard deviation of κ . The cell radius a_s and the distance a_1 of the attachment point from the cell centre are evaluated from experimental images. Green squares: $\kappa = 3.3 \times 10^{-22} \text{ J m}^{-1}$, $L = 13 \mu\text{m}$, $B_x = 8.7 \text{ mT}$, $B_y = 9.3 \text{ mT}$, $a_s = 3.2 \mu\text{m}$, $a_1 = 3.2 \mu\text{m}$. Blue diamonds: $\kappa = 3.3 \times 10^{-22} \text{ J m}^{-1}$, $L = 21 \mu\text{m}$, $B_x = 8.7 \text{ mT}$, $B_y = 9.3 \text{ mT}$, $a_s = 2.7 \mu\text{m}$, $a_1 = 0 \mu\text{m}$. Red circles: $\kappa = 3.3 \times 10^{-22} \text{ J m}^{-1}$, $L = 24 \mu\text{m}$, $B_x = 8.9 \text{ mT}$, $B_y = 10.3 \text{ mT}$, $a_s = 3.1 \mu\text{m}$, $a_1 = 3.1 \mu\text{m}$.

position depends on the field strength since it shifts from $S_p = 1.8$ to 2.8 when the dimensionless field strength M_n varies from 3.4 to 10.3 . The maximum normalized velocity corresponds for each value of the magnetoelastic number to the optimal balance between elasticity, magnetically induced internal forces and viscous forces.

In our experimental realization of a controllable swimming microdevice, a magnetically actuated flexible filament has the ability to exhibit, in a precise regime, a non-reciprocal motion responsible for propulsion. The internal torques, originating from a combination of dipolar colloidal interactions and anisotropy of the susceptibility, can be externally controlled so that the velocity and the direction of motion can be chosen. This swimming device displays a maximum normalized velocity as a function of the dimensionless parameter S_p , as already predicted in the case of model elastic filaments. However, in our device this optimal combination is shown to depend also on a distinct parameter, M_n , that involves the actuating field strength. We hope that a deeper understanding of this effect may clarify the consequences of actuation type and mechanism on the optimal conditions for swimming in nature. These magnetically actuated colloidal devices might possibly also be useful in the precise and selective positioning of micro-objects or the controlled motion of minute quantities of surrounding fluids^{22,23}.

METHODS

Sample preparation. Biotinylated double-stranded DNAs (315 bp) are synthesized using the polymerase chain reaction (PCR) kit (Expand High Fidelity PCR System) provided by Roche Applied Science using biotinylated primers (Eurogentec) and a λ -DNA template. The resulting DNA solution is purified and its concentration is measured by fluorescence (Picogreen).

Human red blood cells ($6 \mu\text{L}$) are first washed three times in $500 \mu\text{L}$ PBS buffer. Then, they are incubated for 30 min in 1 mL of Biotinylated PEG-NHS ($3,400 \text{ Da}$) provided by Nektar Therapeutics in a PBS buffer. After 1 min the sample is centrifuged at $400g$, and the cells are redispersed in $500 \mu\text{L}$ of PBS buffer containing a pluronic surfactant (F-127) at 0.5% by weight.

Superparamagnetic particles of $1 \mu\text{m}$ in diameter ($\sim 10 \times 10^9$ beads mL^{-1}) provided by Dynal (Dynabeads MyOne Streptavidin) with streptavidin grafted onto their surface (7×10^5 streptavidin molecules per particle) are diluted $10\times$

in a PBS buffer with F-127 ($0.5 \text{ wt}\%$). They are washed twice following this procedure before storing them in this buffer. In a $1,000\text{-}\mu\text{L}$ Eppendorf tube we mix: $0.5 \mu\text{L}$ of 315-bp DNA ($1.4 \times 10^{-6} \text{ M}$), $40 \mu\text{L}$ of PBS buffer containing $0.5 \text{ wt}\%$ of F-127, $5 \mu\text{L}$ of the $10\times$ diluted streptavidin beads, and $10 \mu\text{L}$ of the biotinylated red blood cells solution. The Eppendorf tube is placed in a spatially homogeneous magnetic field (30 mT) for 15 min . $10 \mu\text{L}$ of the solution is then added to $40 \mu\text{L}$ of PBS buffer with F-127 ($0.5 \text{ wt}\%$). These assemblies are transferred into a capillary ($2 \text{ cm} \times 100 \mu\text{m} \times 1 \text{ mm}$) that is sealed at both extremities with highly viscous silicone oil (Rhodia). Because this last step can be responsible for the breaking of the structures, we recover them by replacing the capillary in the same magnetic field (30 mT) for 5 min .

Equation of motion. In the Supplementary Information, we derive equations of motion for the filament, which is modelled as a magnetically driven inextensible elastic rod. An expression for the torque resultant upon the filament is given that incorporates both dipole–dipole interactions between the beads in the chain and possible anisotropy in the magnetic susceptibility of the beads. The filament swims close to the floor of the capillary tube, and therefore experiences greatly enhanced drag. The red blood cell is approximately disk-shaped; we determine its radius a_s from the video images, and assume that it has the same approximate thickness (which sets the distance from the filament to the tube floor, h). The red blood cell sits on a thin film of fluid of thickness d , where d is used as a fitting parameter to ensure agreement at high frequency. d is set physically by a combination of steric and electrostatic interactions between the red blood cell glycocalyx and the coating on the capillary tube floor, and by elastohydrodynamic lift forces generated by the moving red blood cell. We suppose that the former effect is dominant, so that d is independent of swimming speed for a given swimmer. The torque and stress resultants vanish at the free end of the filament. At the tethered end, the moment and force are balanced against the rotational and viscous drags upon the red blood cell, which are dominated by the contribution from the lubrication layer.

Parameters measurement. The bending rigidity κ of the filament is measured from the hairpin shape as previously described¹⁷, by using the formula $C_{\text{max}} = \sqrt{\frac{\pi}{3} \frac{B}{2a}} \sqrt{\frac{a^2 \chi^2}{\kappa \mu_0}}$, where C_{max} is the maximum of curvature and χ is the magnetic susceptibility, which is slightly dependent on B and which is given by the manufacturer ($\chi \approx 1$). We find $\kappa = (3.3 \pm 1.6) \times 10^{-22} \text{ J m}^{-1}$. This value is comparable to the one found for real flagella^{24,25}: $\kappa = 4 \times 10^{-22} \text{ J m}^{-1}$. We took $\zeta_{\perp} = 4\pi\eta$ (refs 15, 24 and 26).

Received 25 March; accepted 29 July 2005.

- Bray, D. *Cell Movements: from Molecules to Motility* 6–12 (Garland Publ., New York, 1992).
- Berg, H. C. & Anderson, R. A. Bacteria swim by rotating their flagellar filaments. *Nature* **245**, 380–384 (1973).
- Schuster, S. C. & Khan, S. The bacterial flagellar motor. *Annu. Rev. Biophys. Biomol. Struct.* **23**, 509–539 (1994).
- Gibbons, I. R. Cilia and flagella of eukaryotes. *J. Cell Biol.* **91**, S107–S124 (1981).
- Satir, P. Studies on cilia. 3. Further studies on cilium tip and a sliding filament model of ciliary motility. *J. Cell Biol.* **39**, 77–94 (1968).
- Taylor, G. I. Analysis of the swimming of microscopic organisms. *Proc. R. Soc. Lond. Ser. A* **209**, 447–461 (1951).
- Gray, J. & Hancock, G. J. The propulsion of sea-urchin spermatozoa. *J. Exp. Biol.* **32**, 802–814 (1955).
- Purcell, E. M. Life at low Reynolds number. *Am. J. Phys.* **45**, 3–11 (1977).
- Becker, L. E., Koehler, S. A. & Stone, H. A. On self-propulsion of micro-machines at low Reynolds number: Purcell's three-link swimmer. *J. Fluid Mech.* **490**, 15–35 (2003).
- Najafi, A. & Golestanian, R. Simple swimmer at low Reynolds number: Three linked spheres. *Phys. Rev. E* **69**, 062901–062904 (2004).
- Cox, R. G. Motion of long slender bodies in a viscous fluid. Part I. General theory. *J. Fluid Mech.* **45**, 791–810 (1971).
- Brennen, C. Locomotion of flagellates with mastigonemes. *J. Mechanochem. Cell Motil.* **3**, 207–217 (1976).
- Avron, J. E., Gat, O. & Kenneth, O. Optimal swimming at low Reynolds numbers. *Phys. Rev. Lett.* **93**, 186001–186004 (2004).
- Lagomarsino, M. C., Capuani, F. & Lowe, C. P. A simulation study of the dynamics of a driven filament in an Aristotelian fluid. *J. Theor. Biol.* **224**, 215–224 (2003).
- Wiggins, C. H. & Goldstein, R. E. Flexive and propulsive dynamics of elastica at low Reynolds number. *Phys. Rev. Lett.* **80**, 3879–3882 (1998).
- Lowe, C. P. Dynamics of filaments: modelling the dynamics of driven microfilaments. *Phil. Trans. R. Soc. Lond. B* **358**, 1543–1550 (2003).
- Goubault, C. et al. Flexible magnetic filaments as micromechanical sensors. *Phys. Rev. Lett.* **91**, 260802–260805 (2003).
- Koenig, A. et al. Magnetic force probe for nanoscale biomolecules. *Phys. Rev. Lett.* (in the press).

19. Strick, T. R., Allemand, J. F., Bensimon, D., Bensimon, A. & Croquette, V. The elasticity of a single supercoiled DNA molecule. *Science* **271**, 1835–1837 (1996).
20. Zahn, K., Lenke, R. & Maret, G. Friction coefficient of rod-like chains of spheres at very-low Reynolds-numbers.1. Experiment. *J. Phys. II* **4**, 555–560 (1994).
21. Meunier, A. Friction coefficient of rod-like chains of spheres at very-low Reynolds-numbers.2. Numerical simulations. *J. Phys. II* **4**, 561–566 (1994).
22. Terray, A., Oakey, J. & Marr, D. W. M. Microfluidic control using colloidal devices. *Science* **296**, 1841–1844 (2002).
23. Darnton, N., Turner, L., Breuer, K. & Berg, H. C. Moving fluid with bacterial carpets. *Biophys. J.* **86**, 1863–1870 (2004).
24. Camalet, S., Julicher, F. & Prost, J. Self-organized beating and swimming of internally driven filaments. *Phys. Rev. Lett.* **82**, 1590–1593 (1999).
25. Ishijima, S. & Hiramoto, Y. Flexural rigidity of echinoderm sperm flagella. *Cell Struct. Funct.* **19**, 349–362 (1994).
26. Wiggins, C. H., Riveline, D., Ott, A. & Goldstein, R. E. Trapping and wiggling: Elastohydrodynamics of driven microfilaments. *Biophys. J.* **74**, 1043–1060 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the Imphy Company for providing us with free Mumetal. We also thank A. Ajdari, J. Prost, J.-B. Salmon and D. Weitz for discussions, and C. Gosse, A. Koenig, F. Montel and C. Goubault for help in material preparation.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.D. (remi.dreyfus@espci.fr).

LETTERS

Biomarker evidence for green and purple sulphur bacteria in a stratified Palaeoproterozoic sea

Jochen J. Brocks¹, Gordon D. Love², Roger E. Summons^{2,3}, Andrew H. Knoll⁴, Graham A. Logan⁵ & Stephen A. Bowden⁶

The disappearance of iron formations from the geological record ~1.8 billion years (Gyr) ago was the consequence of rising oxygen levels in the atmosphere starting 2.45–2.32 Gyr ago^{1–3}. It marks the end of a 2.5-Gyr period dominated by anoxic and iron-rich deep oceans. However, despite rising oxygen levels and a concomitant increase in marine sulphate concentration, related to enhanced sulphide oxidation during continental weathering⁴, the chemistry of the oceans in the following mid-Proterozoic interval (~1.8–0.8 Gyr ago) probably did not yet resemble our oxygen-rich modern oceans. Recent data^{5–8} indicate that marine oxygen and sulphate concentrations may have remained well below current levels during this period, with one model indicating that anoxic and sulphidic marine basins were widespread, and perhaps even globally distributed⁴. Here we present hydrocarbon biomarkers (molecular fossils) from a 1.64-Gyr-old basin in northern Australia, revealing the ecological structure of mid-Proterozoic marine communities. The biomarkers signify a marine basin with anoxic, sulphidic, sulphate-poor and permanently stratified deep waters, hostile to eukaryotic algae. Phototrophic purple sulphur bacteria (Chromatiaceae) were detected in the geological record based on the new carotenoid biomarker okenane, and they seem to have co-existed with communities of green sulphur bacteria (Chlorobiaceae). Collectively, the biomarkers support mounting evidence for a long-lasting Proterozoic world in which oxygen levels remained well below modern levels.

The 1.640 ± 0.003-Gyr-old⁹ Barney Creek Formation (BCF) of the McArthur Group, northern Australia, was deposited in a north–south-trending rift west of the present Gulf of Carpentaria over a known area of 25,000 km². The BCF represents a marine succession that accumulated in a quiet, sub-wave-base environment^{5,10–12}. The dominant lithologies are thinly bedded or planar laminated, dolomitic, carbonaceous and pyritic siltstones and shales, with a typical organic carbon content of 0.2–2%, and locally up to 6%¹³. Samples were collected from 12 drill holes (Supplementary Table S1) representing almost the entire north–south axis of the preserved basin. Biomarkers were extracted, fractionated and analysed by gas chromatography–mass spectroscopy in accordance with procedures described in Supplementary Material.

The bitumens of the BCF contain some of the oldest preserved, unambiguously syngenetic biomarkers known so far¹⁴. The distribution of biomarkers in the 48 samples analysed here shows slight variations with lithology and depositional depth but is generally distinct from distributions currently known from all other periods in Earth's history. Figure 1a shows a chromatogram of the most abundant saturated hydrocarbons in a representative dolomitic shale from the Glyde River region. Although signs of biodegradation

are absent, it is dominated (more than 90% by mass) by a 'hump' or unresolved complex mixture. Regular acyclic isoprenoids with 15–20 carbon atoms are abundant and probably mainly derived from the isoprenoid side chain of (bacterio) chlorophylls. Regular acyclic isoprenoids with 21–25 carbon atoms are present in some samples and are attributed to archaea. Detected for the first time in the Precambrian were the C₄₀ carotenoids lycopane, β-carotane and the tentatively identified γ-carotane. Biological precursors of these biomarkers are common in many phototrophic organisms from all environments, but they are typically preserved as hydrocarbon markers under strongly reducing conditions only.

Another unusual feature is the high relative concentration of 3β-methylhopanes (Fig. 1b), triterpanes commonly attributed to microaerophilic proteobacteria, particularly type I methanotrophs¹⁵. The 3β-methylhopane index for the C₃₁ homologue (C₃₁-3β-MHI), defined as the abundance of the C₃₁ methyl hopane relative to the corresponding C₃₀ desmethyl hopane, averages 5.7% in the BCF (*n* = 9; drill core GR-7); modal values observed in typical Neoproterozoic bitumens and oils are much lower (1–2%). The high 3β-MHI indicates a very high abundance and, presumably, a high activity of type I methanotrophs. Similar conditions were observed in Phanerozoic (0.542 Gyr to present) sediments deposited in sulphate-depleted environments¹⁶. In these environments, sulphate concentrations usually significantly less than 0.5 mM allow methanogenic archaea to outcompete sulphate-reducing bacteria for their common substrates H₂ (ref. 17) and acetate¹⁸, leading to enhanced methane generation. Increased levels of methane reaching the oxycline will, in turn, promote enhanced growth of aerobic methanotrophic bacteria. The high 3β-MHI in the BCF might therefore indirectly indicate sulphate concentrations considerably below modern marine levels (28 mM). This interpretation corroborates sulphide isotopic evidence from the overlying Reward Dolomite, the syndepositional shallow-water facies of the BCF. Reward shales contain pyrite with δ³⁴S averaging about +21.8 ± 1.8‰, which is consistent with pyrite precipitation from contemporaneous sea water strongly depleted in sulphate⁵. Isotope data from carbonate-associated sulphate from Australia and North America indicate that marine sulphate concentrations were below 0.5–2.5 mM between 1.7 and 1.3 Gyr ago⁷, and δ³⁴S data on diagenetic pyrites from basins around the world indicate that sulphate depletion might have been a global phenomenon in this time interval^{6,19}. The results of this biomarker study offer independent evidence that sulphate levels in the McArthur basin were significantly lower than in the present ocean.

In contrast to 3β-methylhopanes, the concentration of cyanobacterial 2α-methylhopanes is extremely low (C₃₁-2α-MHI < 1%; Fig. 1b). This attribute distinguishes these ancient bitumens from

¹Research School of Earth Sciences, The Australian National University, Canberra, ACT 0200, Australia. ²Department of Earth, Atmospheric & Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ³Australian Centre for Astrobiology, Building E8C 153, Macquarie University, NSW 2109, Australia. ⁴Department of Organismic & Evolutionary Biology and Department of Earth & Planetary Sciences, Harvard University, Cambridge, Massachusetts 02138, USA. ⁵Geoscience Australia, GPO Box 378, Canberra ACT 2601, Australia. ⁶Department of Geology & Petroleum Geology, University of Aberdeen, Aberdeen AB24 3UE, UK.

other Proterozoic carbonates, in which 2 α -methylhopanes are almost always abundant²⁰. However, not all cyanobacteria biosynthesize 2-methylhopanoids and so the paucity of these biomarkers does not exclude cyanobacteria as an important component of the ecosystem. On the contrary, cyanobacteria are the most likely source for the abundant β -carotane in the BCF (Fig. 1a) because potential precursor pigments are rare in phototrophic sulphur bacteria (PSB),

and eukaryotic algae were not an important component of the ecosystem (see below). Cyanobacteria, together with green non-sulphur bacteria (Chloroflexaceae), are also the most likely source for γ -carotane in the BCF²¹.

In strong contrast to most younger marine environments, biomarker abundances indicate that eukaryotic organisms might have had only an insignificant function in the offshore environment of the

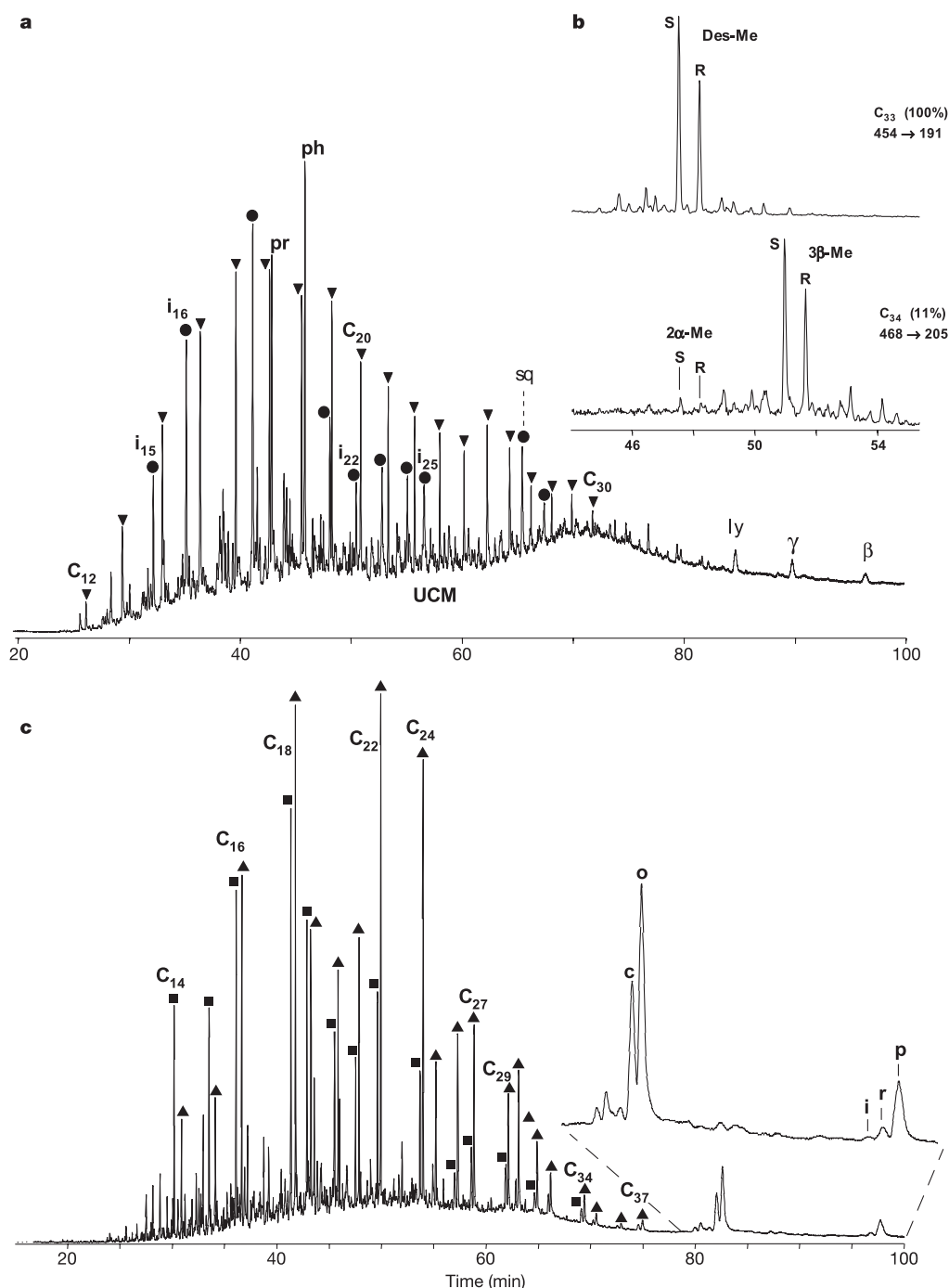


Figure 1 | Biomarker distribution of the BCF. **a**, Total ion recording of the saturated hydrocarbon fraction of sample B03065 (GR-7, 41.80 m). UCM, unresolved complex mixture; triangles, n-alkanes (C_x) with carbon number x ; circles, acyclic isoprenoids (i_x) with carbon number x ; sq, squalane; ly, lycopane; γ , tentatively identified γ -carotane; β , β -carotane. **b**, Multiple reaction monitoring of hopane hydrocarbons in sample B03071 (GR-7, 126.40 m). Chromatograms are identified by carbon numbers, the relative height (abundance) of the most intense peak in the trace, and the reaction

transition from metastable molecular ions fragmenting to daughter ions at mass to charge ratio $m/z = 191$ for desmethylhopanes and $m/z = 205$ for methylhopanes. Top, 22(*S*) and 22(*R*) $\alpha\beta$ -trishomohopane; bottom, their 2 α - and 3 β -methylated analogues. **c**, Selected ion recording at $m/z = 134.1$ of the aromatic fraction of sample B03066 (GR-7, 38.70 m). Squares, 2,3,6-trimethylarylhopanes; triangles, 2,3,4-trimethylarylhopanes; c, chlorobactane; o, okenane (see Fig. 3); i, isorenieratane; r, renieratane; p, renierapurpurane. Structures are shown in Supplementary Information.

BCF. Despite the excellent thermal preservation of the host-rocks, C₂₆ to C₂₉ steranes, biomarkers diagnostic for eukaryotes (and abundant in almost all Phanerozoic bitumens) were close to, or below, the detection limit of 1 p.p.m. In contrast, triaromatic steroids are commonly present at levels of 60–130 p.p.m. However, their distribution is highly unusual (Fig. 2a) and unlike Phanerozoic steroid assemblages (Fig. 2b). The triaromatic steroids entirely lack side-chain alkylation and are predominantly methylated at C-4 (more than 90%), a pattern indicating a possible bacterial source. A plausible biological source of the dominant C-4 methylated steroids is type I methanotrophic bacteria^{22,23}, offering further support for the abundance of these organisms in the BCF. The scarcity of diagnostic eukaryotic steroids in the analysed sample range, despite the abundance of steroids with a typical bacterial distribution, is a phenomenon that has never been observed before in any sedimentary sequence of any age and reflects severely restricted activity of eukaryotic algae in open waters.

The dominant biomarkers (up to 1,000 p.p.m.) in the aromatic hydrocarbon fractions are C₁₄ to C₃₇ 2,3,6- and 2,3,4-trimethyl arylisoprenoids (Fig. 1c), the typical breakdown products of aromatic carotenoids²⁴. In the thermally least altered bitumens, it was possible to detect the intact precursor carotenoids chlorobactane, okenane (Fig. 3), isorenieratane, renieratane and renierapurpurane (chemical structures of the most relevant biomarkers, and original references and additional information about biological sources of aromatic carotenoids, can be found in the Supplementary Information). In living organisms, carotenoids with the renieratane and renierapurpurane skeletons are only known from sponges, in which they might be derived from as yet unknown bacterial sponge symbionts²⁴. There is therefore no reliable biological information for interpreting these biomarkers in the BCF. In contrast, the biological precursors of chlorobactane and isorenieratane, the phototrophic pigments (hydroxy-) chlorobactene and isorenieratene, are well documented²⁴. In deep aquatic environments, such as the BCF,

these carotenoids are derived from planktonic green and brown pigmented species of Chlorobiaceae, respectively. Chlorobiaceae are strictly anaerobic, obligate phototrophs. They require hydrogen sulphide or other reduced sulphur species as an electron donor in the presence of light. Therefore, in planktonic environments, Chlorobiaceae thrive exclusively where euxinic conditions rise into the photic zone of the water column. The only known natural product with the okenane skeleton is okenone (Fig. 3), a red pigment found exclusively in several genera of planktonic purple sulphur bacteria of the family Chromatiaceae²⁵. Although taxonomically unrelated, Chromatiaceae have physiological requirements similar to those of the Chlorobiaceae. Their preferred physiology is phototrophic oxidation of reduced sulphur. Thus, in planktonic environments, they thrive in a thin layer directly below the anoxic–oxic interface within the photic zone.

In modern environments, significant growth of Chromatiaceae occurs only at water depths of up to 20 m but usually much less²⁵. Thus, okenone extracted from recent aquatic sediments most probably indicates an oxycline that ascended (temporarily) to a water depth of ~20 m or shallower. We consider this interpretation also as the most reasonable depth distribution for the source organisms of fossil okenane, assuming that the light requirements of planktonic Chromatiaceae have been stable through time and that planktonic activity and pigment densities in the overlying water column were comparable to those in modern stratified water systems. Similar depth considerations apply to the pigments of Chlorobiaceae. Green pigmented species commonly grow in a thin zone directly below purple sulphur bacteria and are reported at water depths up to ~13 m, whereas the brown pigmented species are adapted to very low light levels²⁵. Brown pigmented green sulphur bacteria are regularly observed at water depths up to 18 m and, in the extreme case of the Black Sea, have been found at 80 m (ref. 26). In modern oceans, an upper mixed layer of less than 20 m is relatively common in warm coastal and even open marine regions²⁷. Marine regions with an

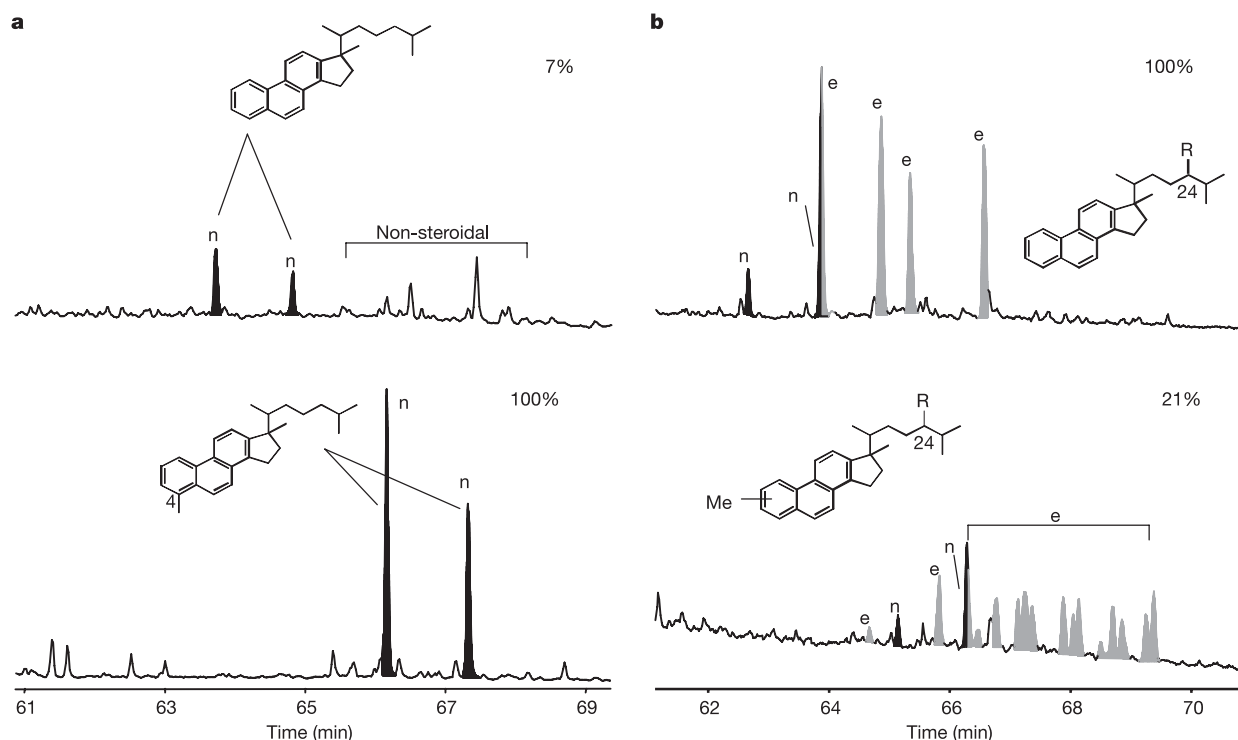


Figure 2 | Selected-ion-recording chromatograms of triaromatic steroids (TA; black and grey signals). Upper traces, $m/z = 231$; lower traces, $m/z = 245$. **a**, A representative sample of the Barney Creek Formation (B03063, GR-7, 180.00 m). **b**, A Phanerozoic petroleum mixture (AGSO Standard).

Unmarked signals are non-steroidal. e (grey signals), diagnostic eukaryotic TA with alkylation at C-24 (R = methyl or ethyl); n (black signals), TA without side-chain alkylation (R = H). Percentages give the relative height (abundance) of the most intense peak in each trace.

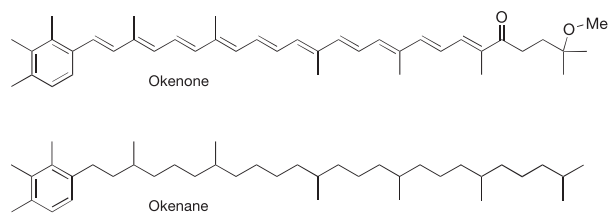


Figure 3 | Chemical structures of the carotenoid pigment okenone and its hydrocarbon fossil equivalent okenane.

upper mixed layer of less than 20 m might also have been widespread in the generally warm mid-Proterozoic, supplying ample marine habitat for anoxygenic phototrophs.

Until now, the oldest known chlorobactane was that reported from the Upper Jurassic²⁸, and we recently detected isorenieratane in the mid-Cambrian Currant Bush Limestone of the Georgina Basin, Australia. Thus, the current findings extend the fossil record of these biomarkers back in time by ~1.5 and ~1.1 Gyr, respectively. Although okenone is found in Recent sediments deposited under euxinic conditions²⁹, the equivalent fossil hydrocarbon okenane has not been reported from sedimentary rocks. It represents the only hydrocarbon biomarker, and in fact the only known fossil indicator, for Chromatiaceae. If okenane and other aromatic carotenoids in the BCF are correctly interpreted as molecular remains of PSB, then the McArthur Basin was intensely stratified and euxinic. The transition from oxic to anoxic conditions occurred just metres to a few tens of metres below the water surface. A shallow mixed layer with low oxygen concentrations and an episodic turbulent influx of H₂S might also explain the suppression of eukaryotic algae, poisoned under these conditions. During periods of widespread euxinia in the oxygen minimum zone, algae would also have suffered nitrogen stress and would have competed poorly against cyanobacteria³⁰. The paucity of diagnostic eukaryotic biomarkers over the analysed range of the basin and over more than 800-m stratigraphic thickness in the Glyde River region indicates that the inferred conditions might have been maintained for extended periods.

Sedimentary successions demonstrating where and how the McArthur Basin was connected to the ocean are not preserved. However, accruing evidence indicates that the anoxic waters of the BCF might have been the extension of an oxygen-deficient ocean adjacent to the north Australian craton. On the eastern periphery of the craton, sulphur isotopes and iron speciation data from successions predating and postdating the BCF and with ages spanning ~1.73–1.49 Gyr indicate that deep-water anoxia and sulphate depletion were a common regional phenomenon^{5,6}. Furthermore, molybdenum isotopes from the same 1.73-Gyr-old and 1.49-Gyr-old sedimentary sequences indicate that, globally, anoxic and dysoxic waters might have covered more of the ocean floor than they do today⁸. The rising evidence for widespread anoxia in mid-Proterozoic marine basins indicates that this was the prime age for planktonic PSB. In the preceding Archaean and early Palaeoproterozoic the supply of reduced sulphur was limited by an excess of dissolved ferrous iron in the oceans, causing rapid precipitation of iron sulphides. With the disappearance of banded iron formations, the availability of sulphide must have increased markedly⁴, supporting blooms of PSB in the McArthur Basin and other marine regions where euxinic waters penetrated the photic zone.

The new biomarker results confirm hypotheses about late Palaeoproterozoic environments and life³⁰. They record a euxinic marine ecosystem with comparatively low sulphate concentrations, high activity of methanotrophic bacteria, blooms of PSB and a paucity of eukaryotic algae in nutrient starved offshore habitats. The oxidation state of surface sea water seems to have risen again beginning at ~1.25 Gyr, perhaps in association with late Mesoproterozoic

(Grenville) orogenesis, but oxygen levels and sulphate concentrations probably only reached more modern values later in the Neoproterozoic³⁰. Further research on biomarkers detected in Proterozoic basins should help to clarify when widespread euxinic conditions retreated and eukaryotic algae expanded into a more nutrient-rich and oxygenated marine habitat.

Received 7 March; accepted 22 July 2005.

- Holland, H. D. in *Treatise on Geochemistry* Vol. 6 (*The Oceans and Marine Geochemistry* (ed. Elderfield, H.)) 583–625 (Elsevier/Pergamon, Oxford, 2004).
- Farquhar, J., Bao, H. & Thiemens, M. Atmospheric influence of Earth's earliest sulfur cycle. *Science* **289**, 756–758 (2000).
- Bekker, A. *et al.* Dating the rise of atmospheric oxygen. *Nature* **427**, 117–120 (2004).
- Canfield, D. E. A new model for Proterozoic ocean chemistry. *Nature* **396**, 450–453 (1998).
- Shen, Y., Canfield, D. E. & Knoll, A. H. Middle Proterozoic ocean chemistry: evidence from the McArthur Basin, northern Australia. *Am. J. Sci.* **302**, 81–109 (2002).
- Shen, Y., Knoll, A. H. & Walter, M. R. Evidence for low sulphate and anoxia in a mid-Proterozoic marine basin. *Nature* **423**, 632–635 (2003).
- Kah, L. C., Lyons, T. W. & Frank, T. D. Low marine sulphate and protracted oxygenation of the Proterozoic biosphere. *Nature* **431**, 834–838 (2004).
- Arnold, G. L., Anbar, A. D., Barling, J. & Lyons, T. W. Molybdenum isotope evidence for widespread anoxia in mid-Proterozoic oceans. *Science* **304**, 87–90 (2004).
- Page, R. W. & Sweet, I. P. Geochronology of basin phases in the western Mt Isa Inlier, and correlation with the McArthur Basin. *Aust. J. Earth Sci.* **45**, 219–232 (1998).
- Bull, S. W. Sedimentology of the Palaeoproterozoic Barney Creek Formation in DDH BMR McArthur 2, southern McArthur Basin, Northern Territory. *Aust. J. Earth Sci.* **45**, 21–31 (1998).
- Jackson, M. J., Southgate, P. N., Winefield, P. R., Barnett, K. & Zeilinger, I. Revised sub-division and regional correlation of the McArthur Basin succession based on NABRE's 1995–8 sequence stratigraphic studies (Australian Geological Survey Organization, record 2000/03, Canberra, 2000).
- Veizer, J., Plumb, K. A., Clayton, R. N., Hinton, R. W. & Grotzinger, J. P. Geochemistry of Precambrian carbonates: V. Late Paleoproterozoic seawater. *Geochim. Cosmochim. Acta* **56**, 2487–2501 (1992).
- Crick, I. H., Boreham, C. J., Cook, A. C. & Powell, T. G. Petroleum geology and geochemistry of Middle Proterozoic McArthur Basin, northern Australia. II: Assessment of source rock potential. *AAPG Bull.* **72**, 1495–1514 (1988).
- Summons, R. E., Powell, T. G. & Boreham, C. J. Petroleum geology and geochemistry of the Middle Proterozoic McArthur Basin, northern Australia. III. Composition of extractable hydrocarbons. *Geochim. Cosmochim. Acta* **52**, 1747–1763 (1988).
- Summons, R. E. & Jahnke, L. L. in *Biological Markers in Sediments and Petroleum* (eds Moldovan, J. M., Albrecht, P. & Philp, R. P.) 182–200 (Prentice Hall, Englewood Cliffs, New Jersey, 1992).
- Collister, J. W., Summons, R. E., Lichtfouse, E. & Hayes, J. M. An isotopic biogeochemical study of the Green River oil shale. *Org. Geochem.* **19**, 265–276 (1992).
- Hoehler, T. M., Alperin, M. J., Albert, D. B. & Martens, C. S. Thermodynamic control on hydrogen concentrations in anoxic sediments. *Geochim. Cosmochim. Acta* **62**, 1745–1756 (1998).
- Scholten, J. C. M. *et al.* Effect of sulfate and nitrate on acetate conversion by anaerobic microorganisms in a freshwater sediment. *FEMS Microbiol. Ecol.* **42**, 375–385 (2002).
- Hayes, J. M., Lambert, I. B. & Strauss, H. in *The Proterozoic Biosphere: A Multidisciplinary Study* (eds Schopf, J. W. & Klein, C.) 129–134 (Cambridge Univ. Press, New York, 1992).
- Summons, R. E., Jahnke, L. L., Hope, J. M. & Logan, G. A. 2-Methylhopanoids as biomarkers for cyanobacterial oxygenic photosynthesis. *Nature* **400**, 554–557 (1999).
- Palmisano, A. C., Cronin, S. E. & Des Marais, D. J. Analysis of lipophilic pigments from a phototrophic microbial mat community by high performance liquid chromatography. *J. Microbiol. Methods* **8**, 209–217 (1988).
- Volkman, J. K. Sterols in microorganisms. *Appl. Microbiol. Biotechnol.* **60**, 496–506 (2003).
- Bird, C. W. *et al.* Steroids and squalene in *Methylococcus capsulatus* grown on methane. *Nature* **230**, 473–474 (1971).
- Brocks, J. J. & Summons, R. E. in *Treatise on Geochemistry* Vol. 8 (*Biogeochemistry*) (ed. Schlesinger, W. H.) 63–115 (Elsevier, Oxford, 2004).
- Van Gernerden, H. & Mas, J. in *Anoxygenic Photosynthetic Bacteria* (eds Blankenship, R. E., Madigan, M. T. & Bauer, C. E.) 49–85 (Kluwer Academic, Dordrecht, 1995).
- Repeta, D. J., Simpson, D. J., Jørgensen, B. B. & Jannasch, H. W. Evidence for the existence of anoxygenic photosynthesis from the distribution of bacteriochlorophylls in the Black Sea. *Nature* **342**, 69–72 (1989).
- Kara, A. B., Rochford, P. A. & Hurlburt, H. E. Mixed layer depth variability over

- the global ocean. *J. Geophys. Res.* **108**(C3), 3079 (2003) (doi:10.1029/2000JC000736).
28. van Kaam-Peters, H. M. E. & Sinninghe Damsté, J. S. Characterization of an extremely organic sulphur-rich, 150 Ma old carbonaceous rock: palaeoenvironmental implications. *Org. Geochem.* **27**, 371–397 (1997).
29. Schaeffer, P., Adam, P., Wehrung, P. & Albrecht, P. Novel aromatic carotenoid derivatives from sulfur photosynthetic bacteria in sediments. *Tetrahedr. Lett.* **38**, 8413–8416 (1997).
30. Anbar, A. D. & Knoll, A. H. Proterozoic ocean chemistry and evolution: a bioinorganic bridge? *Science* **297**, 1137–1142 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Hope, C. Sandison and P. Greenwood for technical support; A. Bradley and R. Haese for scientific expertise on methanogens; Geoscience Australia (GA) and the Northern Territory Geological

Survey for samples; D. Rawlings for expert advice on the geology of the McArthur Group; and P. Schaeffer for supplying synthetic standards of aromatic carotenoids. This work was supported by the William F. Milton Fund of Harvard University and GA. Work conducted at Massachusetts Institute of Technology was supported by a NASA Exobiology grant to R.E.S. G.D.L. and S.A.B. were at the University of Newcastle upon Tyne during parts of the preparation of this work and thank the Natural Environment Research Council for funding a postdoctoral fellowship and PhD studentship, respectively. J.J.B. acknowledges the Harvard Society of Fellows and the Department of Organismic & Evolutionary Biology, Harvard University, for financial support during the preparation of this work.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.J.B. (jochen.brocks@anu.edu.au).

Nonlinear dynamics, granular media and dynamic earthquake triggering

Paul A. Johnson^{1*} & Xiaoping Jia^{2*}

The 1992 magnitude 7.3 Landers earthquake triggered an exceptional number of additional earthquakes within California and as far north as Yellowstone and Montana^{1–3}. Since this observation, other large earthquakes have been shown to induce dynamic triggering at remote distances—for example, after the 1999 magnitude 7.1 Hector Mine¹ and the 2002 magnitude 7.9 Denali⁴ earthquakes—and in the near-field as aftershocks⁵. The physical origin of dynamic triggering, however, remains one of the least understood aspects of earthquake nucleation^{1–5}. The dynamic strain amplitudes from a large earthquake are exceedingly small once the waves have propagated more than several fault radii. For example, a strain wave amplitude of 10^{-6} and wavelength 1 m corresponds to a displacement amplitude of about 10^{-7} m. Here we show that the dynamic, elastic-nonlinear behaviour of fault gouge perturbed by a seismic wave may trigger earthquakes, even with such small strains. We base our hypothesis on recent laboratory dynamic experiments conducted in granular media, a fault gouge surrogate^{6,7}. From these we infer that, if the fault is weak^{8–10}, seismic waves cause the fault core modulus to decrease abruptly and weaken further. If the fault is already near failure, this process could therefore induce fault slip.

Several dynamic triggering mechanisms have been proposed, based primarily on fluid-mechanical interaction in the fault gouge (rock that has been highly fractured and ‘worked’ by the adjacent crustal blocks into a granular state^{2,6,7}), because remote triggering appears more commonly in geothermal areas³. Proposed mechanisms include increased pore pressure associated with the following: the compaction of saturated fault gouge (liquefaction), leading to failure¹; cyclic fatigue of gouge from the oscillatory wave¹¹; and a ‘sub-critical crack growth’ mechanism in which, at crack tips in wet rocks, chemical reactions are accelerated by the wave stresses, leading to failure^{12,13}. It was reported recently that co-seismic release of CO₂ overpressure from a deep source (presumably rare) was responsible for the triggering of activity related to two large events in northern Italy¹⁴. It has also been shown that triggering takes place in regions not associated with geothermal activity^{15–19}. Thus, one or more mechanisms must exist that can explain triggering in dry or both wet and dry conditions. In short, there is much speculation about the mechanism, but experimental and field validation is lacking.

We suspect that dynamic elastic nonlinearity of fault gouge might have a function in triggering because we have observed temporary decrease in modulus (material softening) in the laboratory in a variety of rock types under the influence of wave excitation at seismic strains (10^{-6} to 10^{-4}) (ref. 20). Bearing this in mind, we set out to study material softening in granular media in the laboratory, and to understand its relationship to material weakening, a necessary ingredient in dynamic triggering. A schematic diagram of the experimental setup is shown in Fig. 1. The granular medium is

composed of glass beads of diameter $d = 0.6–0.8$ mm, poured into a duralumin cylinder of diameter $D = 30$ mm and filled to a height of $L = 18.5$ mm. The container is closed with two fitted pistons (piezoelectric transducers) and a normal load corresponding to effective pressure P (effective pressure = confining pressure minus pore pressure, and the pore pressure is 1 atm (0.101 MPa)) ranging from 0.07 to 0.3 MPa is applied to the granular sample across the top piston. Before the acoustic measurements, ten cycles of loading and unloading are performed to consolidate the sample. The volume fraction of glass beads thus obtained is found to be 0.63 ± 0.01 . The bead packing is then held under stress for 12 h so that the healing process of asperity contact between the grains, known as ‘aging’^{7,21}, comes to equilibrium.

In the Earth, the fault core may be impacted by all conceivable wave types. Here we limit our studies to compressional P waves (actually Young-mode waves) to discern whether the general effect of modulus reduction takes place. Wave velocities in the glass bead pack were measured with the application of resonance and travelling wave methods. Because of self-amplification, resonance provides the most sensitive means by which to interrogate elastic nonlinear behaviour quantitatively (see Methods). Figure 2a shows resonance curves in the glass bead pack under 0.11 MPa effective pressure. The graph shows a plot of detected amplitude against frequency at progressively increasing input voltages, normalized by the input voltage. As input voltage is increased, the resonance frequency f_r decreases, corresponding to a decrease in velocity and modulus. The resonance peak

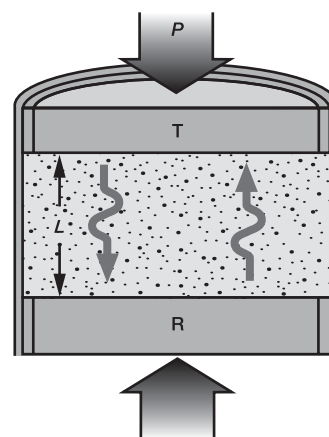


Figure 1 | Diagram of the set-up for conducting resonance and pulse-mode experiments in the glass bead pack under applied pressure P . T and R denote the piezoelectric transmitter and the receiver, respectively, and L is the sample thickness.

¹Geophysics Group EES-11, Los Alamos National Laboratory of the University of California, MS D443, Los Alamos, New Mexico 87545, USA. ²Laboratoire de Physique des Matériaux Divisés et des Interfaces, Université de Marne-la-Vallée, CNRS UMR 8108, Cité Descartes, 77454 Marne la Vallée cedex 2, France.

*These authors contributed equally to this work.

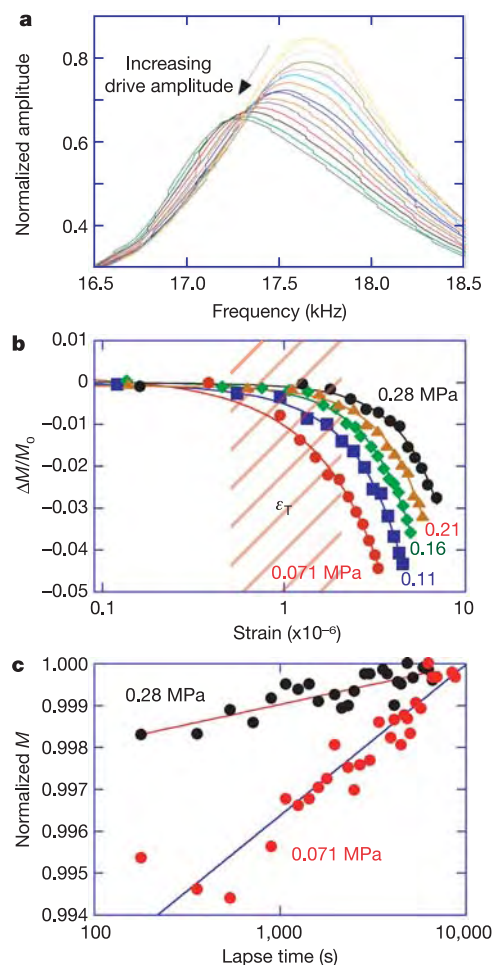


Figure 2 | Material softening due to nonlinear dynamics under resonance conditions. **a**, Resonance curves of the fundamental P-wave mode with increasing input amplitude measured at the detector R (see Methods). **b**, The change in normalized modulus $\Delta M/M_0 = (M - M_0)/M_0$ with detected strain at five effective pressures as noted. M is the modulus as a function of amplitude, and M_0 is the low-amplitude (linear) modulus. **c**, Slow dynamics recovery of the modulus under two different effective pressures, showing the recovery of the modulus with lapse time. The modulus is normalized to its rest, equilibrium value.

broadening and amplitude decrease is an indication of significant, simultaneous nonlinear dissipation²². The change in f_r of about 2.6% corresponds to a decrease in Young modulus of about 5.2% over a strain amplitude range of 10^{-7} to 5×10^{-6} . For comparison, strains measured from seismic waves at hundreds of kilometres from a moderate sized earthquake are of order 10^{-7} to 10^{-6} , depending on the distance travelled and the source radiation pattern^{1,3,4}.

To explore the influence of effective pressure on the nonlinear response, our experimental procedure is repeated at five progressively increasing pressures, as shown in Fig. 2b. Modulus softening diminishes progressively as the pressure is increased, meaning that the system elastic nonlinearity decreases with increasing pressure. The decrease in normalized modulus change $\Delta M/M_0$ ($\sim 2\Delta f/f_r = 2\Delta V/V_0$, where M_0 is the linear, equilibrium modulus) ranges from about 4.8% to 3% between 0.071 and 0.28 MPa effective pressure, respectively, for strains ranging from 10^{-7} to 7×10^{-6} . Thus, in the fault core, we may expect the same behaviour if the fault is weak and/or the effective pressure is low. There is evidence that the effective pressure in some fault cores can be very low^{8,9} from high fluid pressure, or that tectonically induced weakness may exist¹⁰ in others. Furthermore, we see that there exists an approximate strain 'threshold' ε_T below which the granular material behaves as a linear

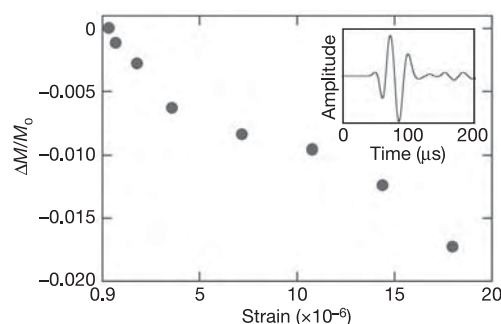


Figure 3 | Relative decrease in modulus with input amplitude in the travelling-wave experiment for the glass-bead pack under an effective pressure of 0.11 MPa. The source signal is a one-cycle sinusoidal pulse at 50 kHz (inset), input at progressively larger input levels from low amplitude (linear regime) to high amplitude (nonlinear regime). The actual wave velocity of about 630 m s^{-1} and modulus obtained in the linear regime (low amplitude) are identical to those obtained in the resonance experiments shown in Fig. 2. The inset shows an example of a detected waveform.

elastic medium (cross-hatched region in Fig. 2b); the resonance frequency and consequently the modulus are independent of strain amplitude below ε_T . ε_T increases progressively with effective pressure but is of order 10^{-6} over the effective pressures studied. We argue below that ε_T is significant in triggering.

We find that the material softening has memory, termed 'slow dynamics'²³, meaning that the modulus slowly returns to equilibrium over several hours or even days after the wave energy has disappeared. Figure 2c shows the recovery of modulus under two different pressures after excitation at large wave strains (about 5×10^{-6}); the plot of $\Delta M/M_0$ against lapse time follows a logarithmic law (although with significant scatter). The dynamic-induced modulus change and successive recovery depend strongly on the effective pressure; for example, $\Delta M/M_0$ after 10^4 seconds (2.8 h) is about 1% under equilibrium for 0.071 MPa and 0.2% under equilibrium for 0.28 MPa (Fig. 2c). The actual duration of the slow dynamics in granular media is as yet only partly quantified because thermal effects in the laboratory that affect material velocity ultimately contaminate the observation, especially late in recovery time. At a minimum, the material is left in a softened state for hours after wave excitation. We argue below that slow dynamics could also have a function in triggering. Logarithmic recovery of slow dynamics has been observed in materials including rock^{22,23} and in the ageing of granular materials that are disturbed by shear motion²⁴ or stirring. However, in the present experiment no visible rearrangement of grains was induced by acoustic vibration. These processes may be due to frictional healing of asperity contact between grains, weakened or broken during the wave vibration⁷.

We now examine the dynamical nonlinear response of the granular material with travelling-wave experiments, analogous to field circumstances in which a seismic wave impinges on a fault. We employed the same experimental set-up as that described in Fig. 1. Figure 3 shows that the modulus softening is immediate on perturbation by the wave pulse and also exhibits dependence on the wave amplitude. In comparison with the resonance studies (Fig. 2b), the dynamically induced reduction in modulus is less pronounced for an equivalent strain amplitude: the magnitude of modulus softening also depends on wave duration. This phenomenon, known as 'conditioning', is observed in other systems including rocks, some ceramics and damaged solids^{22,23}. Here, conditioning more than doubles the modulus change for the equivalent strain amplitude in resonance. Nonetheless, the travelling wave measurements indicate that at a microstrain, elastic nonlinear effects appear and modulus reduction is initiated (always with the caveat that the core is weak). Finally, we note that very recent field experiments in granular media²⁵

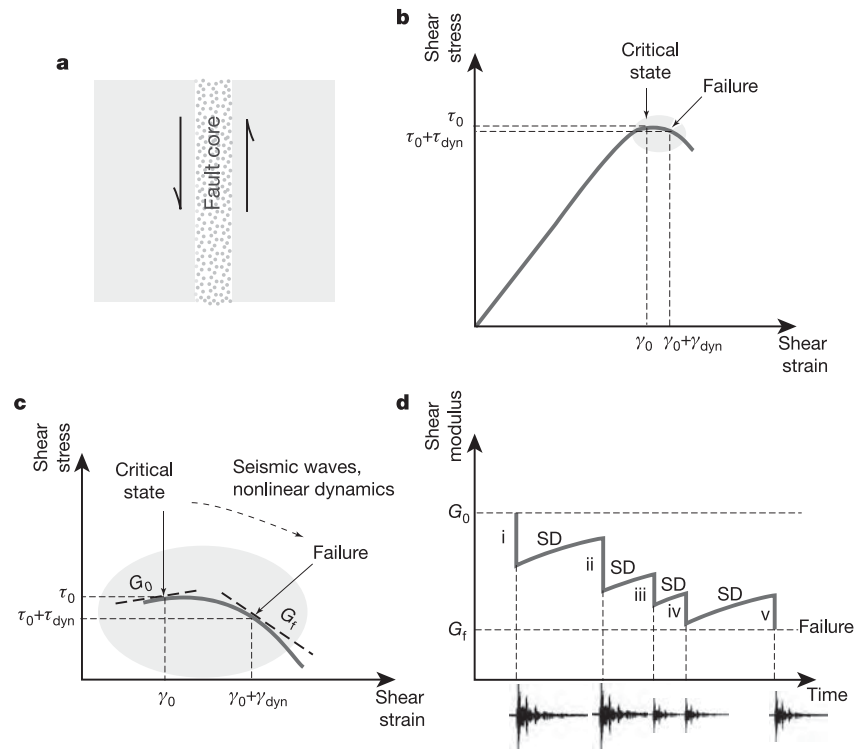


Figure 4 | Failure model: how wave nonlinear dynamics forces a fault system that is in a critical, jammed state to failure by softening and weakening the gouge. **a**, The physical system of a fault core and the surrounding fault blocks. **b**, The shear-stress (τ) versus shear-strain (γ)

fault-core response. **c**, Expanded view of the region of instability noted by the ellipse in **b**. See Methods for definition of variables and details. **d**, The influence of slow dynamics (SD) where successive seismic waves drive the fault core to failure. See Methods for a full description of the process.

at seismic frequencies show all of the behaviours described in the present laboratory experiments.

We believe the physical mechanism responsible for the modulus softening of the granular material is related to the nonlinear frictional properties at the contacts between the grains. A very simple phenomenological model that captures the general nature of the material elasticity can be described as follows. At a given effective stress σ_0 , the dynamic stress σ_{dyn} is^{20,26}

$$\sigma_{\text{dyn}} = M \varepsilon_{\text{dyn}} (1 + \beta \varepsilon_{\text{dyn}} + \delta \varepsilon_{\text{dyn}}^2 + \dots) \quad (1)$$

where ε_{dyn} is the dynamic strain, M is the modulus, and β and δ are the first-order and second-order dynamic nonlinear parameters that describe the shape of the curves in Fig. 2b (a more comprehensive relation is hysteretic but will not affect the proposed model of triggering). It can be inferred from Fig. 2b that β and δ vary with the effective pressure and thus with σ_0 . From contact mechanics²⁷ β and δ are proportional to $1/\varepsilon_0$ and $1/\varepsilon_0^2$, respectively. For $\varepsilon_0 = 1.3 \times 10^{-4}$, corresponding to an applied pressure of about 0.11 MPa, β and δ are order of -7.7×10^5 and -5.9×10^7 , respectively (for comparison, $|\beta| < 10$ for steel). Slow dynamics and conditioning related to the healing process of grain-grain contacts can also be included phenomenologically in this model, in which M has a log(time) dependence.

Taking the above equation of state, we relate the material softening to weakening, by applying logic resembling that of Rice and Rudnicki^{28,29}. We consider the soft fault core surrounded by competent fault blocks in Fig. 4a and the effect of a seismic wave impinging on the system that is in a critical state, near failure (Fig. 4b). To produce significant material softening and simultaneous weakening, taking the fault gouge through the instability and failure (Fig. 4c), the laboratory experiments described here show that the perturbing wave strain must be at least about 10^{-6} . This is shown by the cross-hatched region in Fig. 2b. We suggest that this is why most

seismic waves, even from large earthquakes, do not cause triggering (except in what is traditionally deemed the aftershock zone^{5,30})—their strain amplitudes tend to be 10^{-7} to 10^{-6} at regional distances^{1,3,4}. Only large earthquakes that focus sufficiently large high amplitudes of several microstrains, such as the Denali and Landers events, cause triggering beyond the aftershock zone³⁰. In these cases, the fault core at equilibrium must be in a critical state, in which it can be taken through instability to failure by the perturbation of the seismic wave. Equivalently, we may also interpret triggering as the onset of sliding of the fault, resulting from an abrupt decrease in the shear strength of the granular gouge by break or loss of contact due to strong vibration⁷. Figure 4d describes how slow dynamics could also have a function in inducing delayed triggering in the situation where successive seismic waves impinge on the fault from a foreshock–mainshock–aftershock sequence.

Thus, our laboratory experiments in combination with the softening-to-weakening model presented here indicate that dynamic elastic nonlinearity of the fault core—the gouge—offers an explanation for the occurrence of dynamic triggering in response to seismic waves. Slow dynamics might be responsible for delayed triggering. The necessary physical characteristics for this triggering mechanism require three factors: first, a weak fault (or one with low effective pressure); second, a fault in a critical state; and third, dynamic strain amplitudes greater than about 10^{-6} .

METHODS

Resonance experiment. To optimize the propagation of coherent P waves, large piezoelectric transducers of diameter $D = 30$ mm placed in direct contact with the glass beads are used as the excitation source and receiver²¹ (Fig. 1). The vibration displacement u_{dyn} of piezoelectric transducers is calibrated by an optical interferometer measuring displacement directly on the transducer face, and calculating dynamic strain ε_{dyn} according to $\varepsilon_{\text{dyn}} = du_{\text{dyn}}/dx = 2\pi u_{\text{dyn}}/\lambda$, where λ is the wavelength. We measure resonance in our confined, granular samples over a frequency-sweep interval that contains the fundamental modes.

At each frequency interval in the sweep, the frequency is held fixed until the wave amplitude reaches steady-state conditions. We then extract the time-averaged amplitude at each frequency in the sweep to construct a resonance curve composed of frequency versus amplitude. Curves are obtained at progressively increasing amplitude levels, as seen in Fig. 2a. The resonance frequency shift measured at the curve peak plotted against its detected strain amplitude is then used to characterize the material nonlinearity (Fig. 2b). This is accomplished by monitoring the resonance-induced reduction of effective wave velocity or elastic modulus, according to the relationship $f_r = V/2L$, obtained in a resonator of thickness L with rigid boundary conditions, where f_r is the resonance peak frequency and V the P-wave velocity, and also $V = (M/\rho)^{1/2}$ where M is the Young modulus and ρ is the material density. The granular density of our glass beads stressed by means of a jackscrew arrangement undergoes expansion during the resonance experiments; at 0.11 MPa, for instance, the effective pressure P is observed to increase by about 3%, which would correspond to an increase in modulus of $\Delta M/M_0 \approx 1\%$ according to the hertzian contact elasticity²⁷. This observation, probably arising from a dilatancy-like effect in our dense glass bead packs, implies that the decrease in modulus induced by acoustic vibration might be even stronger than that illustrated in Fig. 2b.

Slow dynamics experiment. In the slow dynamics measurement, the equilibrium, elastically linear modulus of the sample is first measured by applying a low-amplitude strain (of order 10^{-7}) step-sweep and constructing a resonance curve as described above. The amplitude at the resonance peak is recorded. The sample is then vibrated at fixed frequency near the resonance at a large strain amplitude, corresponding to the maximum amplitudes shown in Fig. 2b, for 7 min to induce material softening. Immediately on termination of the high-amplitude excitation, the step-sweep measurement recommences at very low strain to probe the recovery of the resonant peak frequency (modulus) as a function of elapsed time, as shown in Fig. 2c.

Fault core softening, weakening and triggering. The fault core's mechanical response is controlled by the gouge itself and/or roughness on the blocks, presumed granular in nature, shown in Fig. 4a. The stress-strain tensor of the fault core is

$$\varepsilon_{ij} = M_{ijkl}^{-1} \sigma_{kl} \quad (2)$$

The fault core fails in shear, so we consider single shear-stress (τ) and shear-strain (γ) components of σ_{ij} and ε_{ij} , $i \neq j$, respectively. The shear modulus $G = \partial\tau/\partial\gamma$ is assumed much smaller in the core than the surrounding blocks. As shown in Fig. 4b and its zoom (indicated by the ellipse) in Fig. 4c, the fault core is at equilibrium because of the ambient stress field, but in a critical state near failure as denoted by the shear stress and strains τ_0 and γ_0 , where

$$\tau_0 = G_0 \gamma_0 \quad (3)$$

and the core modulus is G_0 . The total stress τ_{tr} is the sum of the equilibrium stress τ_0 and the contribution from the transient seismic wave τ_{dyn} ,

$$\tau_{tr} = \tau_0 + \tau_{dyn} \quad (4)$$

where

$$\tau_{dyn} = G_0 \gamma_{dyn} (1 + \beta \gamma_{dyn} + \delta \gamma_{dyn}^2 + \dots) \quad (5)$$

from equation (1). Substitution of equation (3) and equation (4) into equation (5) yields

$$\tau_{tr} = G_0 \left[\gamma_0 + \gamma_{dyn} (1 + \beta \gamma_{dyn} + \delta \gamma_{dyn}^2 + \dots) \right] \quad (6)$$

If the dynamic strain is zero, the equation reverts to equation (3). Thus, when γ_{dyn} is at a strain amplitude (more than about 10^{-6}) to cause sufficient modulus reduction, equation (6) shows how the core may quickly go from a critical state to failure. The above model differs markedly from the classical Rice–Rudnicki^{28,29} model in that the behaviour of the fault core perturbed by a seismic wave drives the system rather than the static shear stress of the surrounding fault blocks.

Delayed triggering. Figure 4d shows the influence of slow dynamics (SD) where successive seismic waves drive the fault core to failure. In the case where the first wave impinging on the fault core does not cause failure, there is nonetheless a decrease in modulus from equilibrium G_0 denoted by i . The core modulus begins the SD recovery towards equilibrium (SD) but does not reach it. A successive seismic wave drives the modulus even lower (ii) and the SD recommences. With successive seismic waves, $G(t)$ ratchets progressively downwards (iii–v), recovering to some extent between seismic waves, and so on, until one wave drives the system to failure (v). Here the modulus can be described by

$$G(t_i) = G_0 (1 + 2\beta \gamma_{dyn} + 3\delta \gamma_{dyn}^2) \quad (7)$$

where $G(t_i)$ is a function of log(time) after each successive event.

Received 1 February; accepted 11 July 2005.

- Gomberg, J., Reasenber, P. A., Bodin, P. & Harris, R. A. Earthquake triggering by seismic waves following the Landers and Hector Mine earthquakes. *Nature* **411**, 462–466 (2001).
- Scholz, C. H. *The Mechanics of Earthquakes and Faulting* 2nd edn (Cambridge Univ. Press, Cambridge, 2002).
- Hill, D. P. *et al.* Seismicity remotely triggered by the magnitude 7.3 Landers, California, earthquake. *Science* **260**, 1617–1623 (1993).
- Gomberg, J., Bodin, P., Larson, K. & Dragert, H. Earthquake nucleation by transient deformations caused by the $M = 7.9$ Denali, Alaska, earthquake. *Nature* **427**, 621–624 (2004).
- Gomberg, J., Bodin, P. & Reasenber, P. A. Observing earthquakes triggered in the near field by dynamic deformations. *Bull. Seismol. Soc. Am.* **93**, 118–138 (2003).
- Astrom, J. A., Herrmann, H. J. & Timonen, J. Granular packings and fault zones. *Phys. Rev. Lett.* **84**, 638–641 (2000).
- Marone, C. Laboratory-derived friction laws and their application to seismic faulting. *Annu. Rev. Earth Planet. Sci.* **26**, 643–696 (1998).
- Hardebeck, J. L. & Hauksson, E. Role of fluids in faulting inferred from stress field signatures. *Science* **285**, 236–239 (1999).
- Miller, S. A. Fluid-mediated influence of adjacent thrusting on the seismic cycle at Parkfield. *Nature* **382**, 799–802 (1996).
- Hardebeck, J. L. & Hauksson, E. Crustal stress field in southern California and its implications for fault mechanics. *J. Geophys. Res.* **106**, 21859–21882 (2001).
- Gomberg, J., Beeler, N. M., Blanpied, M. L. & Bodin, P. Earthquake triggering by transient and static deformations. *J. Geophys. Res.* **103**, 24411–24426 (1998).
- Brodsky, E. E., Karakostas, V. & Kanamori, H. A new observation of dynamically triggered regional seismicity: Earthquakes in Greece following the August, 1999, Ismir, Turkey Earthquake. *Geophys. Res. Lett.* **27**, 2741–2744 (2000).
- Das, S. & Scholz, C. Off-fault aftershock clusters caused by shear stress increase? *Bull. Seismol. Soc. Am.* **71**, 1669–1675 (1981).
- Miller, S. A. *et al.* Aftershocks driven by high-pressure CO₂ source at depth. *Nature* **427**, 724–727 (2004).
- Gomberg, J. & Bodin, P. Triggering of the Little Skull Mountain, Nevada earthquake with dynamic strains. *Bull. Seismol. Soc. Am.* **84**, 844–853 (1994).
- Gomberg, J., Bodin, P. & Reasenber, P. Observing earthquakes triggered in the near field by dynamic deformations. *Bull. Seismol. Soc. Am.* **93**, 118–138 (2003).
- Hough, S. E. Triggered earthquakes and the 1811–1812 New Madrid, Central United States earthquake sequence. *Bull. Seismol. Soc. Am.* **91**, 1574–1581 (2001).
- Mohamad, R. *et al.* Remote earthquake triggering along the Dead Sea Fault in Syria following the 1995 Gulf of Aqaba earthquake ($M_s = 7.3$). *Seismol. Res. Lett.* **71**, 47–52 (2000).
- Singh, S. K., Anderson, J. G. & Rodriguez, M. Triggered seismicity in the Valley of Mexico from major earthquakes. *Geofis. Int.* **37**, 3–15 (1998).
- Guyot, R. A. & Johnson, P. A. The astonishing case of mesoscopic elastic nonlinearity. *Phys. Today* **52**, 30–35 (1999).
- Jia, X. Coda-like multiple scattering of elastic waves in dense granular media. *Phys. Rev. Lett.* **93**, 154303 (2004).
- Johnson, P. A. & Sutin, A. Slow dynamics and anomalous nonlinear fast dynamics in diverse solids. *J. Acoust. Soc. Am.* **117**, 124–130 (2005).
- TenCate, J. & Shankland, T. J. Slow dynamics in the nonlinear elastic response of Berea sandstone. *Geophys. Res. Lett.* **23**, 3019–3022 (1996).
- Hartley, R. R. & Behringer, R. P. Logarithmic rate dependence of force networks in sheared granular materials. *Nature* **421**, 928–931 (2003).
- Pearce, F. *et al.* Nonlinear soil response induced in situ by an active source at Garner Valley. *Eos* **85** (Suppl.), Abstract S42A–04 (2004).
- Ostrovsky, L. & Johnson, P. A. Dynamic nonlinear elasticity in geomaterials. *Riv. Nuovo Cim.* **24**, 1–46 (2001).
- Johnson, K. L. *Contact Mechanics* (Cambridge Univ. Press, Cambridge, 1985).
- Rudnicki, J. W. The inception of faulting in a rock mass with a weakened zone. *J. Geophys. Res.* **82**, 844–854 (1977).
- Rice, J. & Rudnicki, J. W. Earthquake precursory effects due to pore fluid stabilization of a weakened fault zone. *J. Geophys. Res.* **84**, 2177–2184 (1979).
- Gomberg, J. & Johnson, P. A. Dynamic triggering of earthquakes. *Nature* doi:10.1038/nature04167 (this issue).

Acknowledgements We thank J. Gomberg, C. Marrone, P. Bodin, J. Vidale, A. Freed, R. Guyot, S. Hough, C. Scholz, P. Mills, C. Lewis, S. Meadows and J. Laurent. Funding was provided by the US Department of Energy, Office of Basic Energy Science; CNRS France and the Institute of Geophysics and Planetary Physics at Los Alamos. P.J. also acknowledges a grant from CNRS.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.J. (paj@lanl.gov).

Pterosaur diversity and faunal turnover in Cretaceous terrestrial ecosystems in China

Xiaolin Wang¹, Alexander W. A. Kellner², Zhonghe Zhou¹ & Diogenes de Almeida Campos³

New specimens and an analysis of the Jehol pterosaur faunas of northeastern China show an unexpected diversity of flying reptile groups in terrestrial Cretaceous ecosystems^{1–4}. Here we report two new pterosaurs that are referred to European groups previously unknown in deposits of northeastern China. *Feilongus youngi*, from the Yixian Formation^{1,3}, is closely related to the Gallodactylidae^{5–6} and is distinguished by the presence of two independent sagittal crests and a protruding upper jaw. *Nurhachius ignacibritoi*, from the Jiufotang Formation^{2,3}, has teeth formed by labiolingually compressed triangular crowns, only previously reported in *Istiodactylus latidens*⁷ from England. With these new discoveries, the Jehol pterosaurs show a wide range of groups including both primitive and derived forms that are not matched by any other deposit in the world. The discoveries also document the turnover of pterosaur faunas, with the primitive Anurognathidae and early archaeopterygoids being replaced by derived pterodactyloids. Furthermore, these deposits offer an opportunity to examine the interaction and competition between birds and pterosaurs—it indicates that the avian fauna during the Lower Cretaceous (and possibly most of the Mesozoic) dominated terrestrial, inland regions, whereas pterosaurs were more abundant in coastal areas.

The record of pterosaurs is strongly biased towards ancient coastal environments^{5,8}, and species from terrestrial deposits are rare, showing little diversity⁹. The Jehol Group, comprising the Yixian and Jiufotang formations, is an exception. The myriad of well-preserved fossils from these deposits is known as the Jehol biota, with some taxa—such as fishes (*Peipiaosteus*, *Sinamia*, *Protopsephurus* and *Yanosteus*), the dinosaur *Psittacosaurus* and the birds *Confuciusornis*^{3,4} and *Jeholornis* (undescribed species)—found (at a generic level) in both formations. Recent dating indicates that the main portion of these terrestrial ecosystems existed 125–120 million years ago².

Although pterosaurs have only been reported in the past decade^{10–13}, extensive collecting in the deposits of northeastern China shows that their diversity rivals that of other important lagerstätten, such as the Upper Jurassic Solnhofen limestones⁵ and the Lower Cretaceous Santana Formation¹⁴, providing a unique insight into the pterosaur fauna that lived far within the interior of the continent, away from the ancient coastlines.

Pterosauria Kaup, 1834

Pterodactyloidea Plieninger, 1901

Archaeopterygoides Kellner, 1996

Feilongus youngi gen. et sp. nov.

Etymology. *Feilongus* from the Chinese spelling feilong, meaning flying dragon; *youngi*, in honour of the late Chinese palaeontologist C. C. Young, who described the first pterosaur from China.

Holotype. Skull and mandible deposited at the Institute of Vertebrate

Paleontology and Paleoanthropology (IVPP), Beijing, China (IVPP V-12539; Fig. 1).

Locality and horizon. Jianshangou Bed, lower Yixian Formation at Heitizigou in Beipiao, Liaoning Province, China.

Diagnosis. Large archaeopterygoid pterosaur (wing span ~2.4 m) that can be differentiated from all other members of the Archaeopterygoides by the following unique features: combination of two sagittal cranial crests (a low premaxillary crest positioned in the middle segment of the rostrum, ending well before the anterior margin of the nasoantorbital fenestra, and a short, bony parietal crest); parietal crest with a rounded posterior margin; protruding upper jaw that is about 10% longer than the lower jaw.

The specimen consists of an almost complete skull and mandible that is lying on its right side (Fig. 1a, b). Several cranial elements are unfused, indicating that this specimen probably represents a sub-adult individual at time of death^{14,15}. The length of the skull is between 390–400 mm (tip of premaxilla to squamosal is 390 mm).

The presence of a confluent naris and antorbital fenestra allows the classification of *Feilongus youngi* within the Pterodactyloidea^{5,6,16}. This new species also shows a laterally placed nasal process and a strongly inclined quadrate relative to the ventral margin of the skull, indicating that it is a member of the Archaeopterygoides (*sensu* Kellner⁶), and is therefore compared with other members of this clade. The nasoantorbital fenestra of *Feilongus youngi* occupies 28.7% of the cranial length between the squamosal and the premaxilla, which is larger than in the Ctenochasmatidae (*Pterodaustro* and *Ctenochasma*⁶) and *Gnathosaurus* but smaller than in all other archaeopterygoids (for example, *Pterodactylus* and *Germanodactylus*). The rostrum (that is, the cranial part anterior to the nasoantorbital fenestra) is more elongated than in other pterosaurs with the exception of the Ctenochasmatidae. *Feilongus youngi* has a concave dorsal margin of the skull, a feature that it shares with *Ctenochasma*, *Pterodaustro*, *Gallodactylus* and *Cycnorhamphus* (Fig. 1c).

A low sagittal crest formed by the premaxillae starts at the region corresponding to the ninth tooth and extends posteriorly, finishing well before the anterior margin of the nasoantorbital fenestra (Fig. 1). A distinct rugose area is present on most of the sagittal crest, indicating that it was covered by a soft extension, as has been reported in some other pterosaurs^{14,17,18}. The premaxillary crest in *Feilongus youngi* differs by being positioned in the middle portion of the rostrum, not extending above the nasoantorbital fenestra and by having a very low bony part. A second sagittal crest is observed on the posterior part of the skull, formed mainly by the parietals (Fig. 1c). This structure is thin and shows a rugose surface, indicating that it probably also bore a soft extension. A parietal crest is also observed in *Gallodactylus* and *Cycnorhamphus*, suggesting that they are closely

¹Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, PO Box 643, Beijing 100044, China. ²Paleovertebrate Sector, Department of Geology and Paleontology, Museu Nacional/Universidade Federal do Rio de Janeiro, Quinta da Boa Vista, São Cristóvão, Rio de Janeiro 20940-040, Brazil. ³Museu de Ciências da Terra/Departamento Nacional de Produção Mineral, avenida Pasteur 404, Rio de Janeiro 22290-240, Brazil.

related to *Feilongus*. In *Feilongus*, however, the bony part of the crest is much shorter and has a distinctive rounded posterior end.

The lower jaw is complete, with a total length of 334 mm (tip to craniomandibular articulation is 321 mm). The anterior end is slightly turned upward. Compared to the upper jaw, the lower jaw is about 27 mm shorter, a unique feature among toothed pterosaurs (Fig. 1c). Only the toothless and more derived pterodactyloid *Pteranodon*¹⁵ has the upper jaw longer than the lower jaw, which was achieved independently in both taxa.

The dentition of *Feilongus youngi* consists of elongated, strongly curved needle-shaped teeth that are confined to the anterior third of the skull and mandible (Fig. 1). On the left side, the upper and lower jaws have 18 and 20 alveoli, respectively, indicating that the total

number of teeth is 76, less than in *Gnathosaurus* and the Ctenochasmatidae but more than in *Germanodactylus* and *Cycnorhamphus* (and probably also *Gallodactylus*). In comparison with other archaeopterydactyloids, *Feilongus youngi* had an estimated wing span of around 2.4 m, making it the largest member of this group known so far.

Dsungaripteroidea Young, 1964

Istiodactylidae Howse, Milner & Martill, 2001

Nurhachius ignaciobritoi gen. et sp. nov.

Etymology. *Nurhachius* from Nurhachi, the name of the founder of the Ch'ing Dynasty in China; *ignaciobritoi*, in honour of the late Brazilian palaeontologist Ignacio M. Brito, who has fostered palaeontological studies in Brazil.

Holotype. A partial skeleton deposited at the IVPP (IVPP V-13288; Fig. 2).

Locality and horizon. Jiufotang Formation (Aptian) at Gonggao, Chaoyang, western Liaoning, China.

Diagnosis. An istiodactylid pterodactyloid that can be differentiated from *Istiodactylus latidens* (the only other member of this clade) by the following unique characters: low skull; absence of a suborbital vacuity; jugal with short lacrimal process; teeth labiolingually compressed with triangular roots subequal to or larger than crowns; alveolar margin of the lower jaw slightly bent upward.

This specimen (Fig. 2a) shows a tightly connected scapula–coracoid and an open suture between the extensor tendon process and the first phalanx, representing a subadult individual^{14,15}. The confluent nasoantorbital fenestra and the proportions of several postcranial elements (for example, humerus, metacarpal IV) indicates that *Nurhachius ignaciobritoi* is a member of the Pterodactyloidea^{6,16}. Other features, such as the presence of a notarium (formed by six fused dorsal vertebrae), allow its classification in the Dsungaripteroidea⁶ (Fig. 2a).

The skull of *Nurhachius ignaciobritoi* is visible from the left side and lacks the posterior end, including the braincase (Figs 2a and 3). The preserved part is 315 mm long, with an estimated total length of 330 mm (premaxilla to squamosal is ~320 mm). The most outstanding cranial feature is the nasoantorbital fenestra, which is very long, occupying approximately 58% of the skull length (premaxilla to squamosal). The lacrimal process of the jugal is thin, similar to the members of the Tapejaridae, but is more inclined posteriorly. According to the cranial reconstructions of *Istiodactylus latidens*^{5,7}, this taxon also shows a jugal with a thin and strongly posteriorly inclined lacrimal process, but in *Nurhachius ignaciobritoi* this process is shorter.

The lower jaw of *Nurhachius ignaciobritoi* is observed from the right side (Fig. 2a). It is long (291 mm) and thin, slightly deeper at the posterior region. The dentition comprises 14 teeth on each side of the upper jaw and 13 on the lower jaw, totalling 54 teeth. Except for two tiny anteriorly projected teeth on the tip of the lower jaw, all teeth are labiolingually compressed, with pointed triangular crowns (Fig. 2b). A marked constriction is present at the limit between root and crown. Better preserved ones show sharp anterior and posterior carinae and lack any ridges. This kind of dentition is only observed in *Istiodactylus latidens*^{5,7}, albeit fewer in number (49).

The postcranial skeleton of *Nurhachius ignaciobritoi* shows almost all parts except for some cervical vertebrae, ribs, the tail and the third and fourth wing phalanges. It presents several postcranial features unique to the Pteranodontoidea (*Pteranodon* plus *Istiodactylus* plus Anhangueridae⁶) such as: tall and spike-like neural spines of the mid-cervical vertebrae, scapula shorter than coracoid, and warped deltopectoral crest of the humerus (Fig. 2a). *Nurhachius ignaciobritoi* also shares with *Istiodactylus* and the Anhangueridae a stout scapula with a marked constricted shaft. As in *Anhangueria*¹⁴, metacarpal III of *Nurhachius* articulates with the carpal region whereas metacarpals I–II are reduced, another feature separating istiodactylids and anhanguerids from *Pteranodon*. The phylogenetic analysis presented here shows that *Nurhachius ignaciobritoi* is closely related to *Istiodactylus*, both

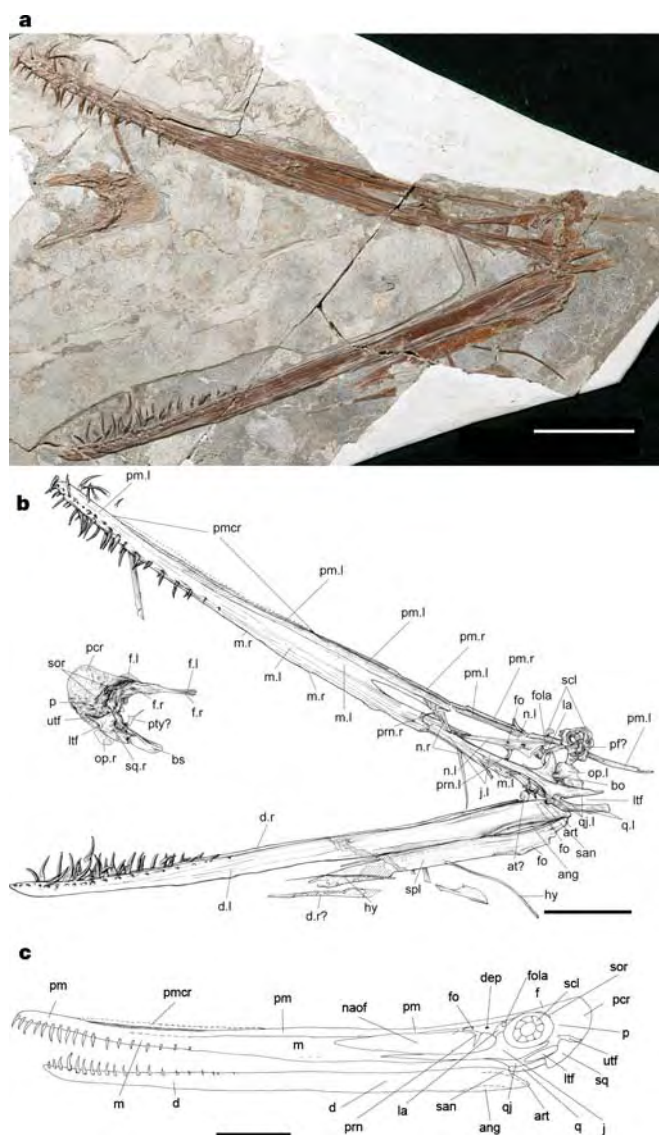


Figure 1 | *Feilongus youngi* gen. et sp. nov., a new archaeopterydactyloid from the Yixian Formation. **a**, Skull. **b**, Drawing showing the contact of cranial bones (the posterior part of the skull is rotated for clarity).

c, Reconstruction. Scale bars, 50 mm. ang, angular; art, articular; at, atlas; bo, basioccipital; bs, basisphenoid; d, dentary; dep, depression; f, frontal; fo, foramen; fola, foramen lacrimale; hy, hyoid bone; j, jugal; la, lacrimal; ltf, lower temporal fenestra; m, maxilla; n, nasal; naof, nasoantorbital fenestra; op, opisthotic; or, orbit; p, parietal; pcr, parietal crest; pf, prefrontal; pl, palatine; pm, premaxilla; pmcr, premaxillary crest; prn, processus nasalis; pty, pterygoid; q, quadrate; qj, quadratojugal; san, surangular; scl, sclerotic ring; sor, supraorbital; spl, splenial; sq, squamosal; utf, upper temporal fenestra. l indicates left and r indicates right.

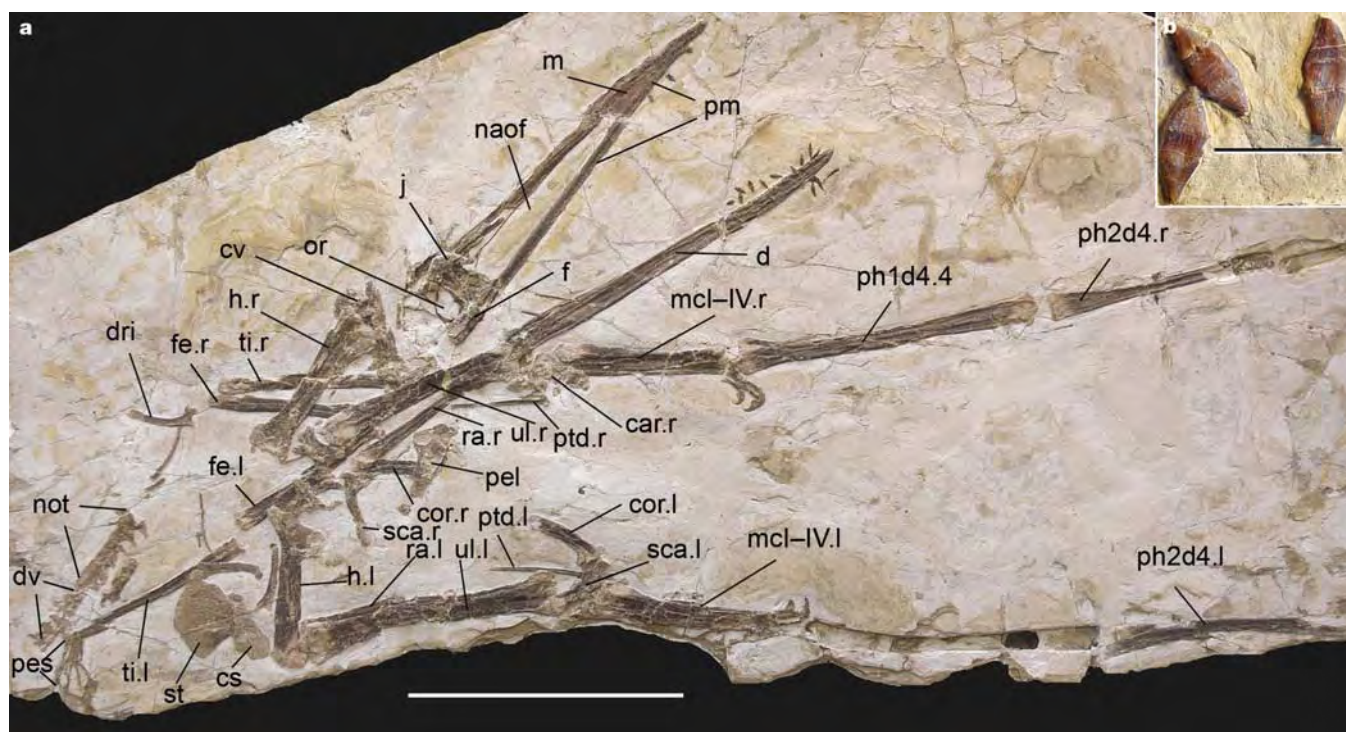


Figure 2 | *Nurhachius ignaciobritoi* gen. et sp. nov., a new istiodactylid from the Jiufotang Formation. a, Complete skeleton. Scale bar, 200 mm. **b**, Detail showing isolated teeth (crown and root). Scale bar, 10 mm. Abbreviations are the same as for Fig. 1 and also include: car, carpus; cor, coracoid; cs, cristospine; cv, cervical vertebra; dri, dorsal rib; dv, free dorsal vertebrae;

fe, femur; hu, humerus; mcl-IV, metacarpal 1-4; not, notarium (fused dorsal vertebrae); pel, pelvis; pes, foot; ph1d4, first phalanx of manual digit IV; ph2d4, second phalanx of manual digit IV; ptd, pteroid; ra, radius; sca, scapula; st, sternum; ti, tibia; ul, ulna.

sharing at least two unique derived features: the particular dentition and the strongly posteriorly oriented lacrimal process of the jugal (see Supplementary Information).

The postcranial elements of *Nurhachius ignaciobrito* indicate a wing span between 2.4 and 2.5 m. Direct comparisons with the European *Istiodactylus latidens* indicate a wing span of 4.2 m for the latter, about 20% less than previously thought⁷.

Including the new taxa, 13 species have been reported from the Jehol Group. The basal Yixian Formation has yielded seven species, including the anurognathids *Dendrorhynchoides curvidentatus*^{10,19} and *Jeholopterus ningchengensis*¹¹, which we show to be more closely related to *Batrachognathus* (from Kazakhstan) than to *Anurognathus* (from Solnhofen limestones; Fig. 4a). *Pterorhynchus* is referred to the Rhamphorhynchidae²⁰ and represents the sole long-tailed pterosaur from the Jehol fauna. The remaining species were referred to the archaeoptero-dactyloid clades Ctenochasmatidae and Pterodactylidae (*Eosipterus yangi*, *Haopterus gracilis* and *Beipiaopterus chenianus*^{10,21,22}) but their exact phylogenetic position has yet to be determined. The case of *Eosipterus*, the first described pterosaur from the Jehol Group, is particularly problematical, because it is still unprepared and several parts were artificially reconstructed (for example, wing elements), casting doubt upon published morphometric studies^{11,19}. Recently, the tapejarid *Sinopterus* was also recorded (specimen IVPP V-14191; X.W., A.W.A.K., Z.Z. & D.d.A.C., unpublished material). *Feilongus youngi* is the first archaeoptero-dactyloid from this deposit to show a close relationship with the Gallodactylidae (*Gallodactylus* plus *Cycnorhamphus*; Fig. 4a).

The overlying Jiufotang Formation has furnished six derived pterodactyloids. Two species (*Sinopterus dongi* and *Sinopterus gui*) are referred to the Tapejaridae^{13,23}, a clade previously known only from the Santana Formation (Brazil²⁴) and Cretaceous deposits of Morocco²⁵. *Liaoningopterus gui* is an anhanguerid¹² and *Chaoyangopterus zhangii* was referred to the Nyctosauridae¹², but is more closely related to *Pteranodon* from the Santonian Niobrara Chalk of

North America. The same is apparently the case for '*Jidapterus edentus*'²⁶, which might be congeneric (or even conspecific) with *Chaoyangopterus*. *Nurhachius ignaciobrito*, described here, is the first istiodactylid from this deposit (Fig. 4).

Despite the recognized difficulties in using pterosaurs in palaeobiogeographic studies, it is interesting to note that the two new taxa—*Feilongus* and *Nurhachius*—have their closest relatives in Europe (Fig. 4). Evidence for broad faunal exchanges between Europe and Siberia/East Asia at this time also comes from other groups, such as iguanodontian ornithopods, dromaeosaurid theropods, enantiornithine birds, discoglossid frogs, paramacellodid lizards, crocodyliforms and gobiconodontid mammals^{3,4,27,28}. Other taxa, such as *Sinopterus* and *Liaoningopterus*, however, have their closest relatives in the Early Cretaceous deposits of Brazil^{12,13,23}, demonstrating that the palaeobiogeographic history of the Jehol biota is very complex⁴.

Despite being two strongly connected deposits (for example, geographical location, age, lithology and depositional environment), the Jehol Group comprises two distinct pterosaur faunas. The Yixian Formation contains the primitive Anurognathidae and

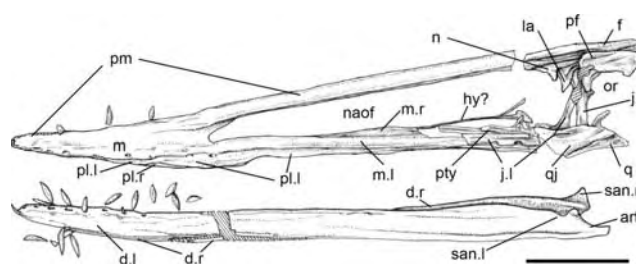


Figure 3 | Drawing of the skull and lower jaw of *Nurhachius ignaciobrito* gen. et sp. nov. Scale bar, 50 mm. Abbreviations are the same as for Fig. 1.

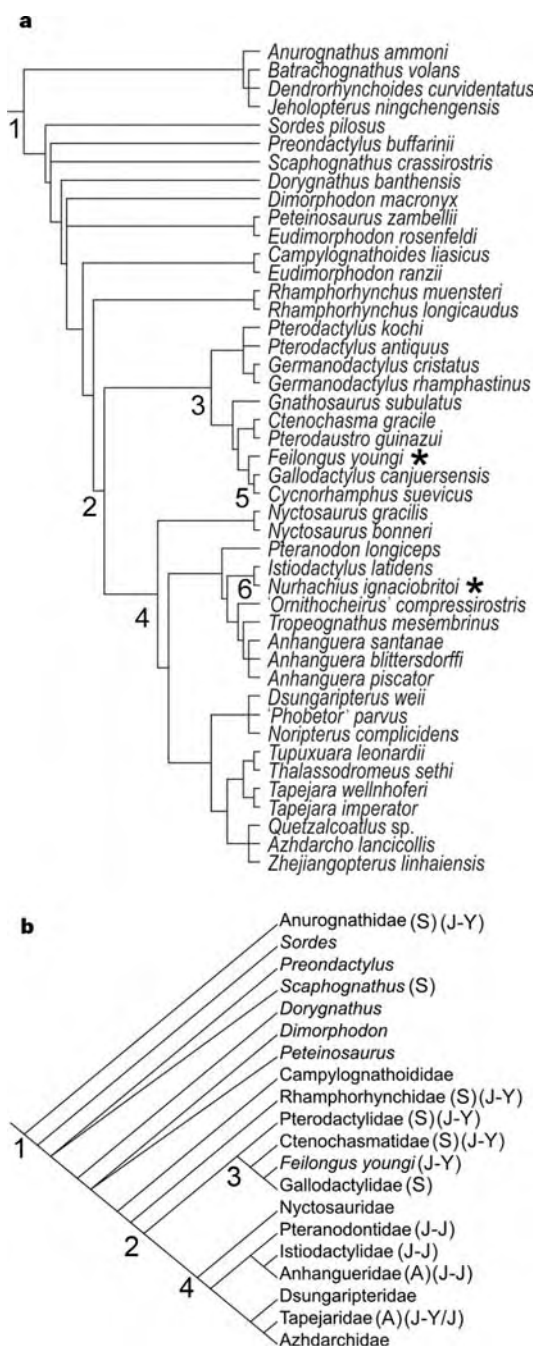


Figure 4 | Cladistic analyses of the new pterosaur species and others from the Jehol Biota. **a**, Cladogram showing the phylogenetic position of *Feilongus youngi* gen. et sp. nov. and *Nurhachius ignaciobrito* gen. et sp. nov. (see Supplementary Information). **b**, Simplified cladogram showing the relationships of the most important pterosaur groups, including several taxa reported from the Jehol Group (based on this study and information from the literature). A, Araripe Basin (Santana Formation, Aptian–Albian); J-J, Jiufotang Formation of the Jehol Group (Barremian–Early Aptian); J-Y, Yixian Formation of the Jehol Group (Barremian–Early Aptian); S, Solnhofen limestones (Early Tithonian); 1, Pterosauria; 2, Pterodactyloidea; 3, Archaeopterodactyloidea; 4, Dsungaripteroidea; 5, Gallodactylidae; 6, Istiodactylidae.

Rhamphorhynchidae, several archaeopterodactyloids and one derived pterodactyloid (the tapejarid *Sinopterus*), which is a very unusual mixture of faunal elements. The Jiufotang Formation comprises derived pterodactyloids such as Istiodactylidae, Tapejaridae, Anhangueridae and possible pteranodontids (*Chaoyangopterus* and

Jidapterus). Combined, the Jehol Group shows a wide range of pterosaur groups (Fig. 4b), a diversity not matched by any other region. The Solnhofen limestones, where fossils have been collected for over two centuries, has yielded primitive forms (Anurognathidae and Rhamphorhynchidae) and members of the Archaeopterodactyloidea, but lacks more derived species⁵. The Santana Formation, where pterosaurs are also found in large numbers, only contains members of the derived Tapejaridae and Anhangueridae (or closely related species^{14,17}). Perhaps the time span encompassing the existence of the Jehol biota—around 5 million years^{1,2}—might explain this diversity, showing a succession from a more archaic pterosaur fauna of the Yixian Formation (with the appearance of some derived forms such as the Tapejaridae) to a more advanced assemblage of the Jiufotang deposits. This time span is greater than the range of the Solnhofen limestones (0.5 million years²⁹) but less than the estimates for the total time of deposition of the Santana Formation (around 8 million years³⁰). It is possible that the Jehol biota documents a worldwide trend in changes of pterosaur faunas, with more primitive forms such as Anurognathidae, Rhamphorhynchidae and early archaeopterodactyloids being replaced by more derived pterodactyloids such as Anhangueridae and probably Pteranodontidae. This unexpected mixture of different pterosaur groups in these Chinese deposits indicates a very complex evolutionary history of pterosaurs in general, which is just beginning to be deciphered.

The Jehol deposits constitute an opportunity to examine the question related to the interaction (and competition) between pterosaurs and birds. The Yixian Formation has furnished an estimated 40 pterosaur remains and more than 1,000 birds. The Jiufotang Formation recorded about 100 pterosaur remains compared with the remains of more than 1,000 birds. Overall, there are 21 avian species described and we know of at least five more. Including the two new pterosaurs described here, there are 13 described and 3 undescribed species. This preliminary analysis clearly shows that birds are more diverse and outnumber pterosaurs (in both the Yixian and Jiufotang formations^{3,4}). This information and comparisons with other deposits^{5,8} leads to the hypothesis that the avian fauna of the Lower Cretaceous—and perhaps most of the Mesozoic era—was more confined to terrestrial, inland regions, whereas pterosaurs dominated the coastal areas. The Jehol deposits are unique Cretaceous terrestrial ecosystems where this question can be refined in the future with more precise stratigraphic information regarding the co-occurrence of these volant creatures.

Received 27 April; accepted 29 June 2005.

- Swisher, C. C. *et al.* A Cretaceous age for the feathered dinosaurs of Liaoning, China. *Nature* **400**, 58–61 (1999).
- He, H. Y. *et al.* Timing of the Jiufotang Formation (Jehol Group) in Liaoning, northeastern China, and its implications. *Geophys. Res. Lett.* **31**, 1–4 (2004).
- Chang, M., Chen, P. J., Wang, Y. Q. & Wang, Y. (eds) *The Jehol Biota* (Shanghai Sci. and Technol. Publishers, Shanghai, 2003).
- Zhou, Z. H., Barrett, P. M. & Hilton, J. An exceptionally preserved Lower Cretaceous ecosystem. *Nature* **421**, 807–814 (2003).
- Wellnhofer, P. *The Illustrated Encyclopedia of Pterosaurs* (Salamander Books, London, 1991).
- Kellner, A. W. A. in *Evolution and Paleobiology of Pterosaurs* (eds Buffetaut, E. & Mazin, J. M.) 105–137 (Geological Society, London, Special Publication 217, 2003).
- Howse, S. C. B., Milner, A. R. & Martill, D. M. in *Dinosaurs of the Isle of Wight* (eds Martill, D. M. & Naish, D.) 324–335 (The Paleontological Association, London, 2001).
- Kellner, A. W. A. Remarks on pterosaur taphonomy and paleoecology. *Acta Geologica Leopoldensia* **39**, 175–189 (1994).
- Kellner, A. W. A. & Langston, W. Jr Cranial remains of *Quetzalcoatlus* (Pterosauria, Azhdarchidae) from the Late Cretaceous Sediments of Big Bend National Park, Texas. *J. Verteb. Paleontol.* **16**, 222–231 (1996).
- Ji, S. A., Ji, Q. & Padian, K. Biostratigraphy of new pterosaurs from China. *Nature* **398**, 573–574 (1999).
- Wang, X. L. *et al.* A nearly complete articulated rhamphorhynchoid pterosaur with exceptionally well-preserved wing membranes and “hairs” from Inner Mongolia, northeast China. *Chinese Sci. Bull.* **47**, 226–230 (2002).

12. Wang, X. L. & Zhou, Z. H. Two new pterodactyloid pterosaurs from the Early Cretaceous Jiufotang Formation of western Liaoning, China. *Vertebrata Palasiatica* **41**, 34–41 (2002).
13. Wang, X. L. & Zhou, Z. H. A new pterosaur (Pterodactyloidea, Tapejaridae) from the Early Cretaceous Jiufotang Formation of western Liaoning, China and its implication for biostratigraphy. *Chinese Sci. Bull.* **48**, 16–23 (2003).
14. Kellner, A. W. A. & Tomida, Y. Description of a new species of Anhangueridae (Pterodactyloidea) with comments on the pterosaur fauna from the Santana Formation (Aptian–Albian), Northeastern Brazil. *Natl Sci. Mus. Monogr.* **17**, 1–135 (2000).
15. Bennett, S. C. The osteology and functional morphology of the Late Cretaceous pterosaur *Pteranodon*. *Palaeontographica Abt. A* **260**, 1–112 (2001).
16. Unwin, D. M. in *Evolution and Paleobiology of Pterosaurs* (eds Buffetaut, E. & Mazin, J. M.) 139–190 (Geological Society, London, Special Publication 217, 2003).
17. Campos, D. A. & Kellner, A. W. A. Short note on the first occurrence of Tapejaridae in the Crato Member (Aptian), Santana Formation, Araripe Basin, Northeast Brazil. *An. Acad. Bras.* **69**, 83–87 (1997).
18. Bennett, S. C. Soft tissue preservation of the cranial crest of the pterosaur *Germanodactylus* from Solnhofen. *J. Vertebr. Paleontol.* **21**, 43–48 (2002).
19. Unwin, D., Lü, J. & Bakhurina, N. N. On the systematic and stratigraphic significance of pterosaurs from the Lower Cretaceous Yixian Formation (Jehol Group) of Liaoning, China. *Mitt. Mus. Nat.kd. Berl. Geowiss. Reihe* **3**, 181–206 (2000).
20. Czerkas, S. A. & Ji, Q. in *Feathered Dinosaurs and the Origin of Flight* (ed. Czerkas, S. J.) 15–41 (The Dinosaur Museum, Utah, 2002).
21. Wang, X. L. & Lü, J. C. The discovery of a pterodactyloid pterosaur from Yixian Formation of western Liaoning, China. *Chinese Sci. Bull.* **46**, 1112–1117 (2001).
22. Lü, J. C. A new pterosaur: *Beipiaopterus chenianus*, gen. et sp. nov. (Reptilia: Pterosauria) from western Liaoning Province of China. *Mem. Fukui Pref. Dino. Mus.* **2**, 153–160 (2003).
23. Li, J. J., Lü, J. C. & Zhang, B. K. A new Lower Cretaceous sinopterid pterosaur from the western Liaoning, China. *Acta Palaeontol. Sin.* **42**, 442–446 (2003).
24. Kellner, A. W. A. & Campos, D. A. The function of the cranial crest and jaws of a unique pterosaur from the Early Cretaceous of Brazil. *Science* **297**, 389–392 (2002).
25. Wellnhofer, P. & Buffetaut, E. Pterosaur remains from the Cretaceous of Morocco. *Paläontologische Zeitschrift* **73**, 133–142 (1999).
26. Dong, Z. M., Sun, Y. W. & Wu, S. Y. On a new pterosaur from the Lower Cretaceous of Chaoyang Basin, Western Liaoning, China. *Global Geol.* **22**, 1–7 (2003).
27. Wu, X. C., Sues, H. D. & Brinkman, D. B. An atoposaurid neosuchian (Archosauria: Crocodyliformes) from the Lower Cretaceous of Inner Mongolia (People's Republic of China). *Can. J. Earth Sci.* **33**, 599–605 (1996).
28. Kielan-Jaworowska, Z., Cifelli, R. L. & Luo, Z. X. (eds) *Mammals from the Age of Dinosaurs—Origins, Evolution, and Structure* (Columbia Univ. Press, New York, 2004).
29. Viohl, G. in *The Beginnings of Birds* (eds Hecht, M. K., Ostrom, J. H., Viohl, G. & Wellnhofer, P.) 31–44 (Freunde des Jura-Museums, Eichstaedt, 1985).
30. Ponte, F. C. & Ponte Filho, F. C. in *Boletim 4 Simpósio Sobre Cretáceo do Brasil* (eds Brito, D. D., Rohn, R. & Perinoto, J. A.) 123–133 (UNESP, Rio Claro, 1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Chang for supporting this work; J. Huang, M. Yang and V. Machado for the drawings and help with preparation of the illustrations; W. Gao for the photos; and Y. Li and L. Xiang for the preparation of the specimens. This study was funded by the National Natural Science Foundation of China, the National 973 Project, Chinese Academy of Sciences, Brazilian Academy of Sciences, Fundação Carlos de Chagas Filho de Amparo a Pesquisa do Rio de Janeiro (FAPERJ) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). A.W.A.K. and D.d.A.C. are members of the Brazilian Academy of Sciences.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to X.W. (xlinwang@263.net) or A.W.A.K. (kellner@acd.ufrj.br).

LETTERS

Trophic cascades across ecosystems

Tiffany M. Knight^{1,2}, Michael W. McCoy¹, Jonathan M. Chase², Krista A. McCoy¹ & Robert D. Holt¹

Predation can be intense, creating strong direct and indirect effects throughout food webs^{1–4}. In addition, ecologists increasingly recognize that fluxes of organisms across ecosystem boundaries can have major consequences for community dynamics^{5,6}. Species with complex life histories often shift habitats during their life cycles⁷ and provide potent conduits coupling ecosystems^{5,6}. Thus, local interactions that affect predator abundance in one ecosystem (for example a larval habitat) may have reverberating effects in another (for example an adult habitat). Here we show that fish indirectly facilitate terrestrial plant reproduction through cascading trophic interactions across ecosystem boundaries. Fish reduce larval dragonfly abundances in ponds, leading to fewer adult dragonflies nearby. Adult dragonflies consume insect pollinators and alter their foraging behaviour. As a result, plants near ponds with fish receive more pollinator visits and are less pollen limited than plants near fish-free ponds. Our results confirm that strong species interactions can reverberate across ecosystems, and emphasize the importance of landscape-level processes in driving local species interactions.

Trophic cascades arise when predators reduce prey abundance, indirectly relaxing consumption on lower trophic levels⁸. For example, in a three-level food chain, predators can reduce herbivore abundance, indirectly benefiting plants. Studies of food-web interactions and trophic cascades have traditionally focused on the antagonistic interactions between species^{1–4,8}. However, mutualists are also embedded within food webs. Many plant species either fail to reproduce or produce fewer seeds or seeds of lower quality if mutualist pollinators, including multitudes of insects, birds and mammals, fail to visit⁹. As a result, pollinators can be crucial drivers of plant population and community dynamics¹⁰. However, pollinators also have important predators; such predators can have an indirect negative effect on plants by harming mutualists^{11–13}.

The strength and ubiquity of trophic cascades has been the focus of a sustained debate in ecology, and considerable effort has focused on quantifying their strength in aquatic and terrestrial ecosystems¹⁴. However, only recently have ecologists explicitly examined how organisms with complex life histories can dynamically couple aquatic and terrestrial ecosystems^{15–17}, leading to trophic cascades that transcend ecosystem boundaries. The larval stages of many freshwater organisms (for example dragonflies and frogs) are vulnerable to a suite of aquatic predators, whereas the adult stages are important consumers in the terrestrial habitat. The intensity of predation experienced by juveniles in the aquatic habitat can therefore be predicted to indirectly influence the intensity of predation imposed in turn by adults in terrestrial habitats.

Here we show how strong direct effects of organisms with complex life histories can create trophic cascades that transcend terrestrial and aquatic ecosystem boundaries. Specifically, we demonstrate that freshwater fish indirectly facilitate plant reproduction by means of a cascade of species interactions, mediated by dragonflies switching during their life history between aquatic and terrestrial habitats (Fig. 1). Fish predation often strongly limits the abundance and

the size distribution (favouring smaller species) of larval odonates (dragonflies and damselflies) in aquatic habitats^{18,19}. Although detailed diet studies on adult dragonflies are rare, published accounts from the general region of our study have shown that the adults of many dragonfly species are voracious predators of bees and other pollinators^{20–22}. We proposed that fish would reduce larval and adult dragonfly abundances and that this would permit a higher abundance of insect pollinators, thus indirectly increasing the pollination and reproductive success of nearby terrestrial plants.

Our study took place at the University of Florida's Katharine Ordway Preserve/Carl Swisher Memorial Sanctuary in northern Florida. This site contains 18 permanent ponds (retain standing water in most years) that differ in whether or not they contain fish. We chose eight ponds; four contained a community of fish (such as Centrarchid sunfishes) and four lacked fish. We found no systematic differences between ponds in surface area or in the amount of sunlight or vegetation structure near the pond margins (Supplementary Information).

As predicted, larval (Fig. 2a) and adult (Fig. 2b) dragonflies were much more abundant in and around fish-free ponds than at ponds with fish. The species composition of dragonflies also differed between ponds with and without fish; large and medium-sized dragonflies dominated in and around fish-free ponds, whereas small species were more prevalent in and around ponds with fish (Fig. 2).

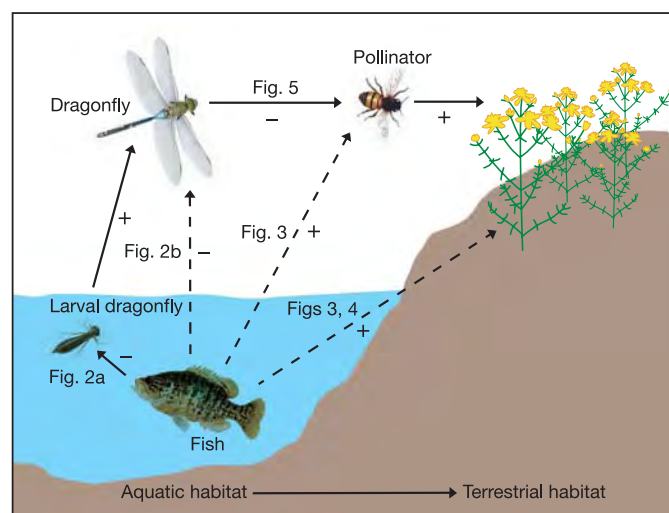


Figure 1 | Interaction web showing the pathway by which fish facilitate plant reproduction. Solid arrows indicate direct interactions; dashed arrows denote indirect interactions. The sign refers to the expected direction of the direct or indirect effect (see the text). Figure numbers indicate which figure presents data supporting each of the predicted effects. (Figure created by S. White and C. Stierwalt.)

¹Department of Zoology, University of Florida, Gainesville, Florida 32611, USA. ²Department of Biology, Washington University, St Louis, Missouri 63130, USA.

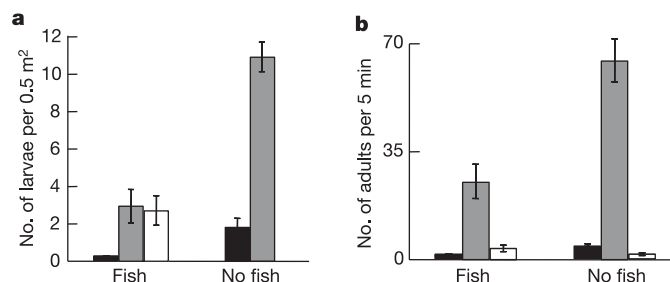


Figure 2 | Surveyed dragonfly abundances in and near fish-containing and fish-free ponds. We sorted dragonfly species into three size categories, small, medium and large (see Methods for dragonfly genera in each category). **a**, There were more larval individuals from the medium (grey bars) and large (black bars) size classes of dragonfly species, and fewer individuals from the small size classes (white bars), in ponds without fish than in ponds with fish (MANOVA: Pillai trace = 0.99, $F_{3,4} = 19.96$, $P < 0.01$; univariate F -tests for large species, $F_{1,6} = 93.12$, $P < 0.001$; for medium species, $F_{1,6} = 85.67$, $P < 0.002$; and for small species, $F_{1,6} = 114.42$, $P < 0.001$). **b**, The abundance of adult dragonflies was lower near ponds with fish (ANOVA: $F_{1,6} = 10.85$, $P < 0.02$). There was a difference between the abundances of adult medium and large dragonfly species near fish-free ponds and near ponds with fish (MANOVA: Pillai trace = 0.82, $F_{3,4} = 5.89$, $P = 0.06$; univariate F -tests for large species, $F_{1,6} = 22.97$, $P < 0.003$; and for medium species, $F_{1,6} = 7.50$, $P < 0.03$); there was no difference in small dragonfly density between fish-containing and fish-free ponds ($F_{1,6} = 1.52$, $P > 0.26$). Results are shown as means $\pm$ s.e.m.

To assess the correlation of fish presence with pollinator visitation, we observed pollinator visitation on one of the most common flowering shoreline plants at these ponds, *Hypericum fasciculatum* (Hypericaceae; St John's wort). Pollinator visitation rates were much higher on *H. fasciculatum* shrubs near ponds with fish than near fish-free ponds (Fig. 3), and there was a difference in the composition of pollinator species; most visitors near ponds with fish were hymenopterans (mostly bees), whereas most visitors near fish-free ponds were dipterans (flies) (Fig. 3). Hypericaceae have evolved traits that attract bees²³; bees may therefore be more effective than flies at pollinating *Hypericum*. Thus, the effect of reduced pollinator visits near fish-free ponds might be magnified, because those few visits that did occur were primarily from less effective pollinators.

We examined whether the effects of fish on dragonflies (Fig. 2) and their cascading effects on pollinator visits (Fig. 3) indirectly

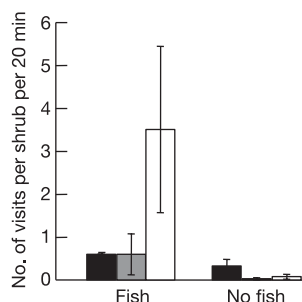


Figure 3 | Pollinator visitation rates to *Hypericum fasciculatum*. The total number of pollinator visits to *Hypericum fasciculatum* was higher near ponds with fish (ANOVA: $F_{1,6} = 11.45$, $P < 0.02$). There was a marginally significant difference between the compositions of pollinators at fish-containing and fish-free ponds (MANOVA, Pillai trace = 0.79, $F_{3,4} = 4.9$, $P = 0.07$). The number of visits by all three groups of pollinators (black bars, Diptera; grey bars, Lepidoptera; white bars, Hymenoptera) was less near ponds with fish (univariate F -tests: Diptera (primarily Syrphidae, Bombyliidae) $F_{1,6} = 4.62$, $P < 0.07$; Hymenoptera (primarily *Agapostemon* spp. (Halictidae)) $F_{1,6} = 15.50$, $P < 0.05$; Lepidoptera (primarily Noctuidae) $F_{1,6} = 5.72$, $P = 0.05$). Results are shown as means $\pm$ s.e.m.

influenced plant reproductive output. We performed pollen supplementation experiments to determine the degree to which *H. fasciculatum* seed production was limited by pollen receipt at each pond. Plants near fish-free ponds were more than twice as pollen limited than plants near ponds with fish (Fig. 4). Complementary experiments with a second plant species (*Sagittaria latifolia*) also showed enhanced pollinator visitation and reproductive output at the margins of ponds with fish (Supplementary Information), indicating that the indirect effect of fish upon plant reproductive success might be general.

Our results indicate that fish presence might have led to low larval and adult dragonfly abundances in and around ponds, causing cascading indirect effects on pollinator visitation rates and plant reproductive output in an adjacent terrestrial community. The decrease in adult dragonfly numbers near ponds with fish probably reflects a combination of demographic effects, in which fish predation reduces larval dragonfly density^{18,19}, and behavioural effects, in which dragonflies avoid ovipositing in ponds with fish. However, because our study did not experimentally manipulate fish presence we cannot discern the relative magnitudes of the possible mechanisms.

We found evidence that pollinator visitation was lower near fish-free ponds, both because adult dragonflies predate on pollinators and because pollinators behaviourally avoid foraging near adult dragonflies. At our study site, over a seven-day period, we observed several predation events by two common species of dragonflies (*Anax junius* and *Erythemis simplicicollis*)²⁴ known to attack large insect species including pollinators; four of eight of those observed predation events were on pollinators (bees, moths and flies; Fig. 5a). To examine the behavioural influence of dragonfly presence on pollinator visitation, we put cages around naturally occurring *H. fasciculatum* near a pond with fish; the mesh size allowed free access by most pollinators but precluded escape by enclosed *E. simplicicollis*. We found that fewer visitors entered cages containing dragonflies than control cages (paired t -test: $t = -4.2$, $P = 0.002$), and visitors that did enter cages with dragonflies foraged on fewer flowers than visitors that entered cages not containing dragonflies ($t = -3.8$, $P = 0.009$). This resulted in *H. fasciculatum* flowers receiving fewer overall visits in the presence of a dragonfly (Fig. 5b).

Strong linkages between consumers in aquatic and terrestrial ecosystems are not limited to this special case in which an aquatic consumer (fish) affects terrestrial predators of mutualists. Many terrestrial predators, herbivores and pollinators have larval aquatic phases. Aquatic predators might therefore have a variety of consequences for interactions in neighbouring terrestrial ecosystems. Similarly, many organisms (for example salamanders) with terrestrial life-stages are important aquatic predators, and thus interactions in

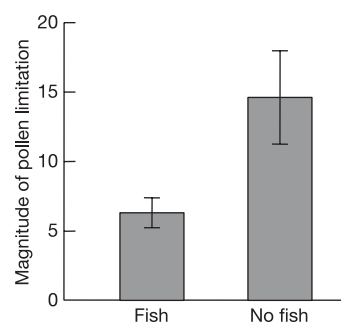


Figure 4 | Results from pollen supplementation experiments. *Hypericum fasciculatum* plants near fish-containing ponds had significantly less pollen limitation (quantified as the difference in the average seed set between experimental supplementation and control treatments) than in plants near fish-free ponds (ANOVA: $F_{1,6} = 7.91$, $P < 0.03$). Results are shown as means $\pm$ s.e.m.

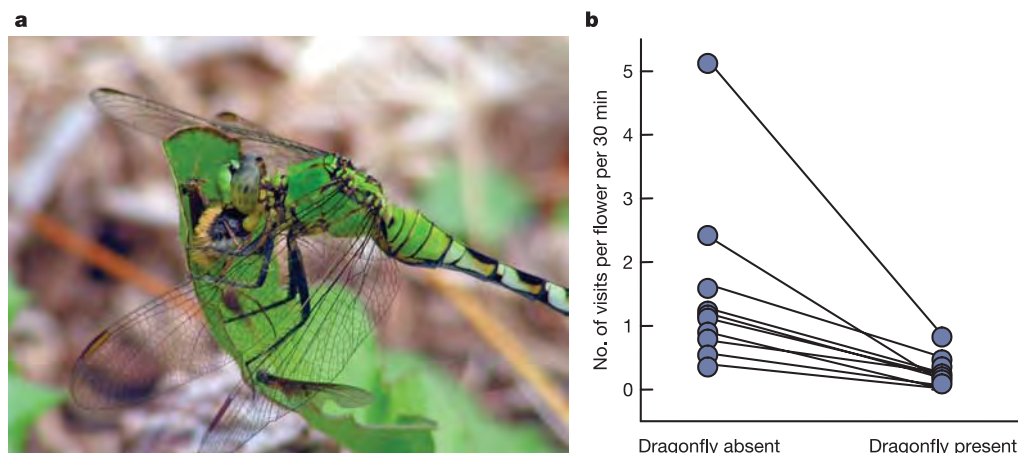


Figure 5 | Effects of dragonflies on pollinators. **a**, Photograph of a dragonfly (female *Erythemis simplicicollis*) consuming a bee-fly pollinator (*Bombylius* sp. (Diptera: Bombyliidae)) at our study site. Photo by M.W.McC. **b**, Results from the experiment comparing pollinator visitation

rates in pairs of large-mesh cages placed around a *H. fasciculatum* shrub, one with a dragonfly (female *E. simplicicollis*) in it, and one as a control. Overall, visitation was much lower in the cage with the dragonfly than in the control cage (paired *t*-test: $t = -3.5$, $P = 0.007$).

the terrestrial ecosystem can cascade to the aquatic ecosystem. Determining the relative strengths of these interactions across ecosystem boundaries should be the focus of future empirical research.

Although the importance of habitat connectance for metacommunity structure has been well studied, this has primarily been restricted to cases involving patches of a single ecosystem type²⁵. Understanding interactions across ecosystem types may be crucial for gauging the effects of anthropogenic environmental change. Deliberate introductions of fish by humans are ubiquitous throughout the world (for example for recreation or pest control)^{26,27}. Our results reveal that such introductions might have cascading effects on adjacent terrestrial ecosystems. By increasing the reproductive success of insect-pollinated plants, freshwater fish introductions potentially alter competitive relationships between terrestrial plants, putting plants not pollinated by insects at a competitive disadvantage. Wetland destruction can harm dragonfly populations, with similar consequences for terrestrial plants. Conversely, a decline in fish abundances (and consequent increase in dragonflies) due to eutrophication, hydroperiod modification or pollution could indirectly harm insect-pollinated plants. Our findings emphasize how consumer flows across radically disparate ecosystems can affect landscape-level processes and drive local interactions between species.

METHODS

Study site and choice of ponds. In May 2003 we chose eight ponds within the Katharine Ordway Preserve/Swisher Memorial Sanctuary (managed by the University of Florida), in Putnam County, Florida. The average distance between study ponds was about 1,000 m (range 200–2,500 m). Four ponds were fish-free; four had abundant fish species (M.W.McC., unpublished observations), including several members of the Centrarchidae, which are known predators of larval odonates¹⁸. Fish-containing and fish-free ponds were interspersed throughout this site and did not differ in surface area (*t*-test: $t = 1.46$; d.f. = 6; $P > 0.20$) or in the structural features of their surrounding vegetation (Supplementary Information).

Quantifying larval and adult dragonflies. We estimated larval dragonfly densities by box sampling¹⁸. In each pond, a 0.5-m² metal box was deployed at five haphazardly chosen locations, and we determined the number of dragonfly larvae present by sweeping a 0.45 m × 0.25 m net until no additional dragonfly larvae were captured in five consecutive sweeps. We quantified adult dragonfly abundances by means of point counts, for which a single observer (M.W.McC.) counted the number of each species of dragonfly that entered the field of view during a 5-min observation period. At each pond, two counts were made from points each located 90° from the other.

We examined adult and larval dragonfly responses to fish presence in two

ways (all data were square-root transformed before analyses to fit normality assumptions). First, we compared the total number of dragonflies between fish-containing and fish-free ponds with ANOVA. Second, to determine whether there were any compositional differences in response to fish, we separated dragonfly species into size classes (based on ref. 28). Genera in the small size class included *Erythrodiplex* and *Celithemus* (Libellulidae); the medium size class included *Libellula*, *Erythemis*, *Pachydiplax* and *Tramea* (Libellulidae); the large size class included *Anax* and *Coryphaesnae* (Aeshnidae). Here we first used MANOVA to determine whether there was an overall treatment effect. After a significant ($P < 0.05$) or marginally significant ($P < 0.08$) MANOVA, we examined univariate *F*-tests to discern differences for each size class.

Pollinator visitation rates. Over a two-week period from late May to early June, we observed pollinators at ten similar-sized *Hypericum fasciculatum* shrubs at each pond. We watched each focal shrub for 20 min and calculated the average number of pollinator visits per shrub per 20 min. First, we compared the visitation rate of plants near fish ponds with those of plants near fish-free ponds by using ANOVA. Second, we separated pollinators into their orders (Hymenoptera, Diptera and Lepidoptera) and used MANOVA to determine whether there was an overall treatment effect, and univariate *F*-tests to discern differences between pollinating taxa.

Pollen supplementation experiments. At each pond we chose ten *H. fasciculatum* shrubs that were similar in size and floral display and separated by at least 10 m from any other study plant. On each plant we chose two similar-sized branches and randomly selected one for the pollen supplement treatment and the other as a control. Flowers in the supplement treatment were rubbed with the anthers of flowers from another plant that was more than 5 m away. Branches were visited daily for one week until most (more than 60%) of the flowers had received pollen supplementation. For each shrub, we calculated the magnitude of pollen limitation as the difference in average seed set (counted under a dissecting microscope) between the supplemented and control flowers; we averaged those values to calculate the magnitude of pollen limitation for each pond. We used ANOVA to compare pollen limitation of shrubs near fish-containing and fish-free ponds.

Dragonfly predation observations. In April 2005, as we walked through the vegetation at the margins of each pond and flushed a variety of insects, we witnessed several predation events by two of the most common species of dragonflies at our study site, *Anax junius* and *Erythemis simplicicollis*. After we witnessed a predation event, we approached the dragonfly while it was consuming its prey and identified the prey to taxa (usually to order) by using close-focusing (about 5 m) binoculars.

Effects of dragonfly presence on pollinator visitation. We constructed cylindrical cages 1.5 m in diameter and 2 m tall from diamond-shaped, plastic mesh fencing 2.3 cm in width, designed to contain adult dragonflies (which cannot retract their wings), but to allow free access by most pollinators (except large lepidopterans, which were almost never observed visiting *H. fasciculatum*). At a single pond with fish, we paired two *H. fasciculatum* shrubs that were similar in size and floral display, and placed one of these cages over each shrub. We then introduced one adult female *E. simplicicollis* into one cage chosen at random and left the other as a caged control. This dragonfly was common at the study site,

especially at ponds without fish, and readily attacks large prey. Further, *E. simplicicollis* is a sit-and-wait predator, and its normal behaviour was not greatly altered by being caged. After 10 min, allowing time for the dragonfly to settle down, one observer watched each caged shrub and recorded the number and identity of pollinators that entered the cage and visited at least one flower, and the number of flowers each pollinator visited on a shrub before leaving. Each shrub was watched for 30 min. We then moved the cages to two new paired shrubs and repeated the experiment with a different dragonfly. In all we performed ten such paired experiments over a four-day period. We performed these experiments between 08:00 and 11:00 on mostly sunny days, so that pollinator activity rates would be comparable between paired experiments. We analysed these data with a paired *t*-test.

Received 14 June; accepted 30 June 2005.

1. Terborgh, J. L. *et al.* Ecological meltdown in predator-free forest fragments. *Science* **294**, 1923–1926 (2001).
2. Worm, B., Lotze, H. K., Hillebrand, H. & Sommer, U. Consumer versus resource control of species diversity and ecosystem functioning. *Nature* **417**, 848–851 (2002).
3. Sinclair, A. R. E., Mduma, S. & Brashares, J. S. Patterns of predation in a diverse predator-prey system. *Nature* **425**, 288–290 (2003).
4. Finke, D. L. & Denno, R. F. Predator diversity dampens trophic cascades. *Nature* **429**, 407–410 (2004).
5. Polis, G. A., Anderson, W. B. & Holt, R. D. Towards an integration of landscape and food web ecology: the dynamics of spatially subsidized food webs. *Annu. Rev. Ecol. Syst.* **28**, 289–316 (1997).
6. Polis, G. A., Power, M. E. & Huxel, G. R. *Food Webs at the Landscape Level* (Univ. of Chicago Press, Chicago, 2004).
7. Werner, E. E. & Gilliam, J. F. The ontogenetic niche and species interactions in size-structured populations. *Annu. Rev. Ecol. Syst.* **15**, 393–425 (1984).
8. Pace, M. L., Cole, J. J., Carpenter, S. R. & Kitchell, J. F. Trophic cascades revealed in diverse ecosystems. *Trends Ecol. Evol.* **14**, 483–488 (1999).
9. Knight, T. M. *et al.* Pollen limitation of plant reproduction: ecological and evolutionary causes and consequences. *Annu. Rev. Ecol. Syst.* **36**, 467–497 (2005).
10. Kremen, C. & Ricketts, T. H. Global perspectives on pollination disruptions. *Conserv. Biol.* **14**, 1226–1228 (2000).
11. Dukas, R. & Morse, D. H. Crab spiders affect flower visitation by bees. *Oikos* **101**, 157–163 (2003).
12. Suttle, K. B. Pollinators as mediators of top-down effects on plants. *Ecol. Lett.* **6**, 688–694 (2003).
13. Muñoz, A. A. & Arroyo, M. T. K. Negative impacts of a vertebrate predator on insect pollinator visitation and seed output in *Chuquiraga oppositifolia*, a high Andean shrub. *Oecologia* **138**, 66–73 (2004).
14. Shurin, J. B. *et al.* A cross-ecosystem comparison of the strength of trophic cascades. *Ecol. Lett.* **5**, 785–791 (2002).
15. Nakano, S. & Murakami, M. Reciprocal subsidies: Dynamic interdependence between terrestrial and aquatic food webs. *Proc. Natl Acad. Sci. USA* **98**, 166–170 (2001).
16. Murakami, M. & Nakano, S. Indirect effect of aquatic insect emergence on a terrestrial insect population through predation by birds. *Ecol. Lett.* **5**, 333–337 (2002).
17. Sabo, J. L. & Power, M. E. River-watershed exchange: Effects of riverine subsidies on riparian lizards and their terrestrial prey. *Ecology* **83**, 1860–1869 (2002).
18. Morin, P. J. The impact of fish exclusion on the abundance and species composition of larval odonates: results of short-term experiments in a North Carolina farm pond. *Ecology* **65**, 53–60 (1984).
19. Johansson, F. & Brodin, T. Effects of fish predators and abiotic factors on dragonfly community structure. *J. Freshwat. Ecol.* **18**, 415–423 (2003).
20. Byers, C. F. *A Contribution to the Knowledge of Florida Odonata* (Univ. of Florida Press, Gainesville, Florida, 1930).
21. Wright, M. Some random observations on dragonfly habits with notes on their predaceousness on bees. *J. Tennessee Acad. Sci.* **19**, 295–301 (1944).
22. Needham, J. G. Notes on some dragonflies of southwest peninsular Florida. *Bull. Brooklyn Entomol. Soc.* **40**, 104–110 (1945).
23. Takhtajan, A. *Diversity and Classification of Flowering Plants* (Columbia Univ. Press, New York, 1997).
24. Corbet, P. S. *Dragonflies: Behavior and Ecology of Odonata* (Cornell Univ. Press, Ithaca, New York, 1999).
25. Leibold, M. A. *et al.* The metacommunity concept: a framework for multi-scale community ecology. *Ecol. Lett.* **7**, 601–613 (2004).
26. Knapp, R. A., Matthews, K. R. & Sarnelle, O. Resistance and resilience of alpine lake fauna to fish introductions. *Ecol. Monogr.* **71**, 401–421 (2001).
27. Moyle, P. B. in *Ecology of Biological Invasions of North America and Hawaii* (eds Mooney, H. A. & Drake, J. A.) 27–43 (Springer, New York, 1986).
28. Dunkle, S. W. *Dragonflies of the Florida Peninsula, Bermuda, and the Bahamas* (Scientific Publishers, Gainesville, Florida, 1989).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T.-L. Ashman, M. Barfield, J. Drake, L. Harmon, J. Knouft, J. Steets, J. M. Chase's lab group and the C. W. Osenberg, B. Bolker and C. St Mary lab group for discussion and comments, T. Kim for field assistance, and S. Coates for logistical support. Funding was provided by The University of Florida Foundation and Washington University in St Louis. Additional support to T.M.K. was through a fellowship from the National Center for Ecological Analysis and Synthesis during the preparation of this manuscript.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.M.K. (tknight@biology2.wustl.edu).

LETTERS

Pathogenic fungus harbours endosymbiotic bacteria for toxin production

Laila P. Partida-Martinez¹ & Christian Hertweck¹

A number of plant pathogenic fungi belonging to the genus *Rhizopus* are infamous for causing rice seedling blight. This plant disease is typically initiated by an abnormal swelling of the seedling roots without any sign of infection by the pathogen^{1–4}. This characteristic symptom is in fact caused by the macrocyclic polyketide metabolite rhizoxin that has been isolated from cultures of *Rhizopus* sp.^{5,6}. The phytotoxin exerts its destructive effect by binding to rice β -tubulin, which results in inhibition of mitosis and cell cycle arrest^{7,8}. Owing to its remarkably strong antimitotic activity in most eukaryotic cells, including various human cancer cell lines, rhizoxin has attracted considerable interest as a potential antitumour drug^{9,10}. Here we show that rhizoxin is not biosynthesized by the fungus itself, but by endosymbiotic, that is, intracellular living, bacteria of the genus *Burkholderia*. Our unexpected findings unveil a remarkably complex symbiotic-pathogenic relationship that extends the fungus–plant interaction to a third, bacterial, key-player, and opens new perspectives for pest control.

Understanding the molecular basis of the biosynthesis of rhizoxin (1 in Fig. 1) is of high importance, not only because of the agricultural implications and its potential use as anticancer agent^{9,10} but also because of the common use of *Rhizopus* species in biotechnology and the production of fermented foods¹¹. Stable isotope labelling studies¹² suggested that rhizoxin biosynthesis involves a polyketide synthase (PKS). Genes encoding these enzymes, in particular their ketosynthase (KS) domains, can be targeted in the genome of bacteria and fungi by polymerase chain reaction (PCR) using specific degenerate primers^{13,14}. In the course of our studies on the molecular basis of rhizoxin biosynthesis, we were, unexpectedly, unable to detect any fungal polyketide synthase genes in the genome of rhizoxin-producing *Rhizopus microsporus*. In contrast, oligonucleotides specifically designed for bacterial type I PKS genes yielded PCR products of the expected size (0.7 kilobases, kb). Sequencing and database searches revealed that the amplified gene fragments are most similar to a class of type I PKS genes encoding so-called ‘trans-AT’ PKS, which are exclusively known from bacteria¹⁵.

This unexpected result could imply that either *Rhizopus* sp. have acquired bacterial PKS genes by horizontal gene transfer, or, alternatively, that the PKS genes were amplified from an associated bacterial microorganism. These considerations would be in accord with the previous report¹⁶ on the isolation of structurally related metabolites from a marine *Pseudomonas* sp. A viable approach to investigate the presence of bacteria is the identification of gene sequences from the bacterial small subunit (16S) ribosomal RNA (rRNA), which is distinct from the small subunit rRNA in eukaryotes. In fact, using universal primers, we were able to amplify bacterial 16S rDNA by PCR from the fungal metagenome. Furthermore, we could rule out a casual association of bacteria by comparison of four rhizoxin producers with two non-rhizoxin-producing

Rhizopus sp. strains from distant geographic areas (Japan, China, Ukraine, Australia; see Table 1).

Both *Rhizopus* species incapable of producing rhizoxin did not give any PCR product with universal primers. Conversely, we could unequivocally prove the presence of bacteria in all rhizoxin producers (Fig. 2a). The entire 16S rDNA (about 1.5 kb) sequence was amplified and sequenced for analysis of the phylogenetic affiliation. First, within a given fungal species, only identical 16S rDNA sequences were obtained, indicating that the populations contained a single bacterial type. Second, database searches revealed that the bacteria belong to the genus *Burkholderia*. Third, on the basis of 16S rDNA from different symbioses, a strongly supported clade was obtained, which contained all bacteria that are associated with different rhizoxin-producers (Fig. 2c). These results clearly demonstrate that rhizoxin-producing *Rhizopus* sp. are persistently associated with a single type of closely related *Burkholderia* sp. and form a symbiosis in the sense of persistent living together of dissimilar organisms. Furthermore, by PCR using KS primers, type I PKS genes were exclusively detected in the metagenome of rhizoxin producers (Fig. 2b), and the analysed gene fragments were highly similar. Thus, all symbionts bear candidate genes for a polyketide synthase that is likely to be involved in rhizoxin biosynthesis.

Although we were able to show the correlation between bacterial association and rhizoxin formation, it was still necessary to establish whether the bacteria occur extracellularly or whether they are true endosymbionts. A co-culture of bacteria and fungi could already be excluded by light microscopic analyses. To visualize the symbiont in the cytosol, fungal preparations from a growing culture of *R. microsporus* ATCC 62417 were stained with a mixture of dyes (SYTO 9 and propidium iodide) that is specific for bacteria and can distinguish between live and dead ones. The samples were treated with a fluorescence enhancer and analysed using a laser microscope at 480 nm/500 nm. As shown in Fig. 3, mycelium of *R. microsporus* harboured a high number of endobacteria that fluoresced as green, rod-shaped spots, indicating that they were alive. To correlate

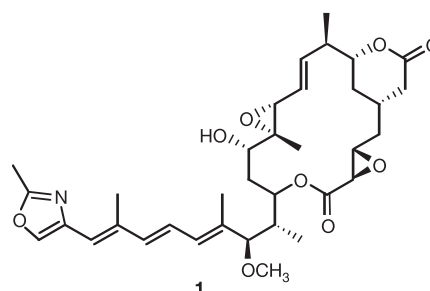


Figure 1 | Structure of rhizoxin, the causal agent of rice seedling blight, isolated from *Rhizopus* sp.

¹Leibniz Institute for Natural Products Research and Infection Biology, HKI, Beutenbergstr. 11a, 07745 Jena, Germany.

Table 1 | *Rhizopus* sp. analysed in this study

Species	Strain	Origin	Rhizoxin producer?	Associated <i>Burkholderia</i> ?	PKS genes detectable?
<i>Rhizopus microsporus</i> van Tieghem var. <i>chinensis</i>	ATCC 62417	Japan (rice seedlings)	Yes	Yes	Yes
<i>Rhizopus</i> sp. F-1360	ATCC 20577	Japan (soil)	Yes	Yes	Yes
<i>Rhizopus microsporus</i> var. <i>microsporus</i>	CBS 699.68 (ATCC 52813)	Ukraine (soil)	Yes	Yes	Yes
<i>Rhizopus microsporus</i> var. <i>microsporus</i>	CBS 308.87	Australia (human tissue)	Yes	Yes	Yes
<i>Rhizopus microsporus</i> van Tieghem var. <i>chinensis</i>	DSMZ 1834	Not specified	No	No	No
<i>Rhizopus microsporus</i> var. <i>chinensis</i> (Saito) Schipper & Stalpers	CBS 631.82	China (bread)	No	No	No

rhizoxin biosynthesis with the presence of bacteria, we aimed at eliminating the symbionts of *R. microsporus* strain ATCC 62417. For this purpose, *Rhizopus* was cultivated in the presence of various antibiotics and the metabolite pattern was monitored by high-performance liquid chromatography-mass spectrometry (HPLC-MS). Out of a number of antibacterial substances that have been found to be active against *Burkholderia* sp.¹⁷, we found that ciprofloxacin completely suppresses rhizoxin production at 40 µg ml⁻¹. HPLC and MS analyses of the fungal culture extract unequivocally revealed that the symbiont-free strain does not produce any detectable amounts of rhizoxin (Fig. 4). Although the

fungus itself did not show any obvious morphological changes, PCR using universal primers as well as KS primers confirmed the presence of a symbiont-free *Rhizopus* strain. This result not only provides an important insight into the question of the true source of the metabolite, but is certainly of great interest for controlling the production of the mycotoxin and pathogenicity factor in agriculture and food biotechnology.

In analogy to Koch's postulates in classical microbiology, the final proof of the *Rhizopus* symbiont hypothesis would be the isolation of the symbiont in pure culture and the re-introduction into a symbiont-free host¹⁸. In order to isolate bacteria associated

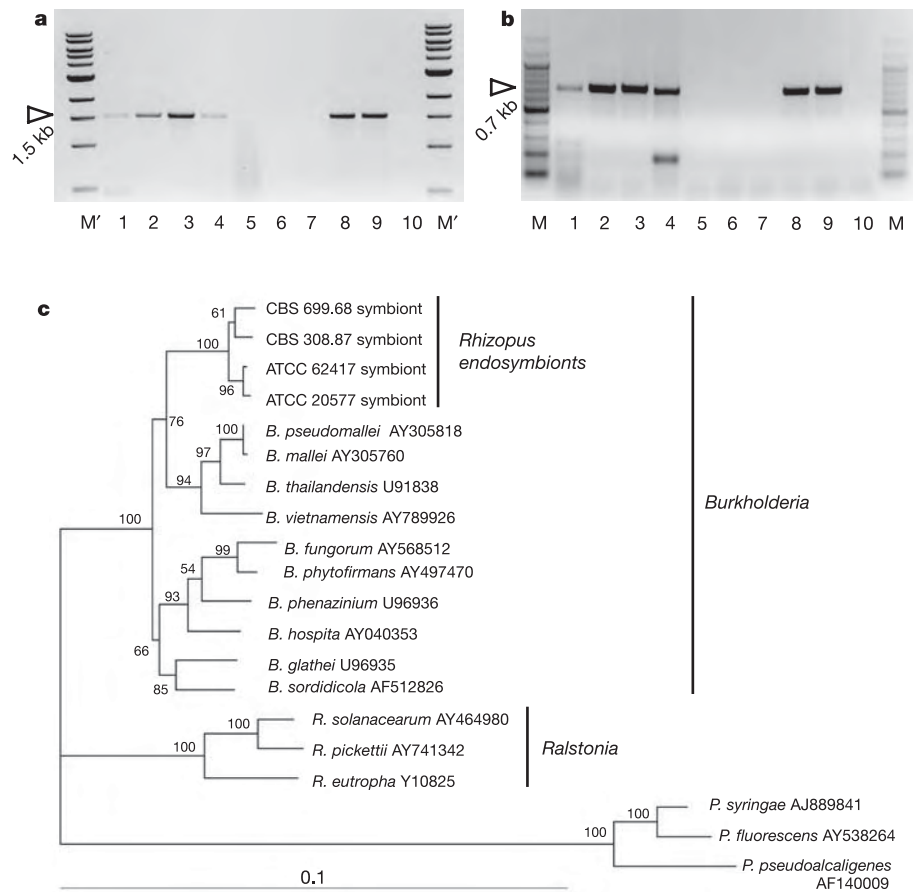


Figure 2 | Results of PCR experiments and phylogenetic affiliation of the symbionts. **a, b,** Agarose (2%) gel electrophoresis of **a**, 16S rDNA and **b**, KS PCR products amplified with universal or KS primers, respectively (see arrow). Fungal metagenomic or bacterial DNA served as templates. Rhizoxin-producing *R. microsporus* strains ATCC 20577, ATCC 62417, CBS 308.87, CBS 699.68 (lanes 1–4); non-producing strains DSMZ 1834, CBS 631.82 (lanes 5, 6); symbiont-free *R. microsporus* ATCC 62417 (lane 7);

genomic DNA of cultivated *Burkholderia* sp. (lane 8); positive and negative controls (lanes 9, 10); 100 bp and 1 kb fragment size markers (M and M'). **c**, Phylogenetic tree based on 16S rRNA of *Rhizopus* endosymbionts and representatives of related *Burkholderia* and *Ralstonia* sp. Sequences were aligned by using the ClustalX program²⁸. The branch comprising species in the genus *Pseudomonas* was used as an outgroup.

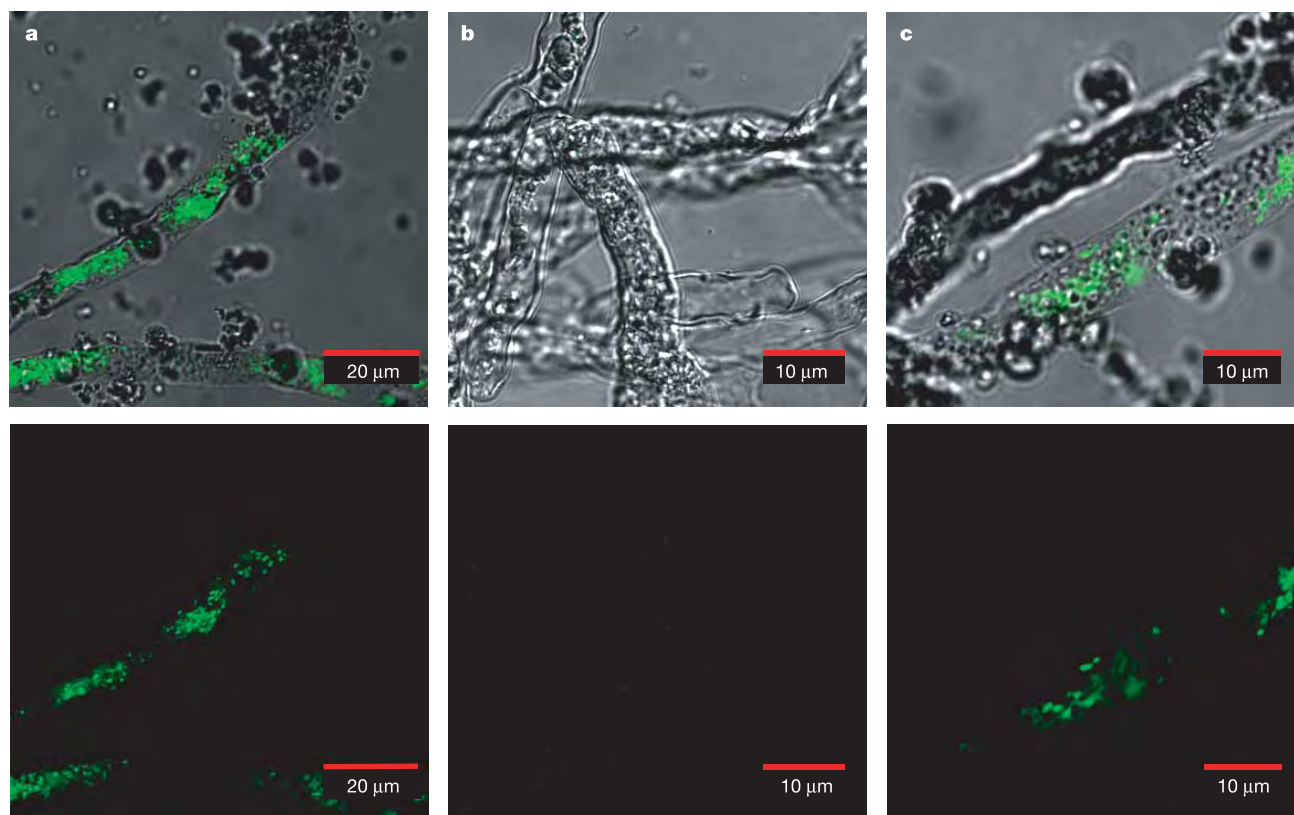


Figure 3 | Laser microscopic investigation of fungal mycelium. **a**, Bacterial endosymbionts in the cytoplasm of *R. microsporus* ATCC 62417; **b**, symbiont-free *R. microsporus* ATCC 62417; **c**, *R. microsporus* ATCC 62417 after reinfection. Mycelium was stained with SYTO[®] 9 from the Live/Dead BacLight Bacterial Viability Kit (Molecular Probes – L7007) and observed

using a Zeiss LSM 510 Meta Laser Microscope at 480/500 nm (top row: white light and fluorescence mode; bottom row: fluorescence mode only). Green fluorescence indicates that bacteria are alive. Scale bars are shown red for clarity.

with *Rhizopus* sp., mycelium from a growing culture was mechanically sheared by pipetting and subjected to gradient centrifugation. After each centrifugation step, aliquots of the supernatant were plated onto nutrient agar plates and incubated. With this procedure, we obtained single colonies of a very slow growing bacterial isolate. PCR and microscopic analyses revealed the presence of a pure culture of rod-shaped motile bacteria. 16S rDNA analyses using bacterial genomic DNA proved the identity of the isolate with the associated bacterial strain. Furthermore, amplified KS gene fragments from the bacterial genome were identical with those obtained from the fungal metagenome. For fermentation of the isolate, a variety of cultivation conditions were tested and the fermentation broths were extracted and analysed. Strikingly, the *Burkholderia* strain produces rhizoxin even in pure culture, as shown by HPLC and MS (Fig. 4, and Supplementary Information). Besides rhizoxin, the symbiont also produces several rhizoxin derivatives (for example, 2,3-deoxyrhizoxin, seco-2,3-deoxyrhizoxin, and nor-2,3-deoxyrhizoxin) that have previously been isolated from the fungus⁶. We also noticed that the isolated symbiont loses the ability to produce rhizoxin in pure culture over time, which may be due to genomic instability or downregulation of biosynthesis genes. However, we observed stable rhizoxin production after re-introduction of the symbiont into the symbiont-free *Rhizopus* strain, which is in full accord with Koch's postulates (Fig. 3 and Supplementary Information). It is likely that chemical signals in the fungal–bacterial association trigger and/or maintain rhizoxin biosynthesis in the state of symbiosis.

In many cases, it has been suspected that natural products isolated from various eukaryotes—for example, tunicates, sponges or insects—are biosynthesized by associated bacterial symbionts and that the animals are not the true producers of the natural product¹⁹.

Only recently, Piel and co-workers demonstrated the existence of symbiotic secondary metabolite producers in *Paederus* beetles and in the sponge *Theonella swinhoei* by striking genomic evidence^{20,21}. Our work provides direct evidence for the endosymbiont hypothesis by generating a symbiont-free *Rhizopus* strain, as well as the isolation and re-introduction of the associated *Burkholderia* strain. Related species belonging to the genus *Burkholderia* are known for their

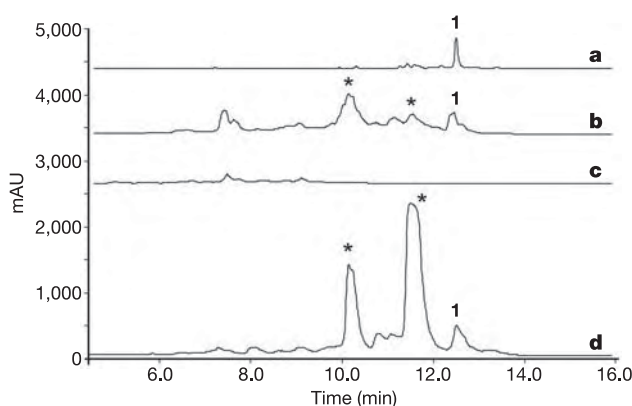


Figure 4 | HPLC profiles of culture extracts monitored at 310 nm for the detection of rhizoxin. **a**, Rhizoxin reference (12.5 mg l^{-1}); **b**, *R. microsporus* ATCC 62417; **c**, symbiont-free *R. microsporus* ATCC 62417 (treated with ciprofloxacin); **d**, *Burkholderia* sp. isolated from *R. microsporus* ATCC 62417. The rhizoxin peak is marked with '1', asterisks indicate peaks due to rhizoxin derivatives. mAU, milli absorbance units.

potential to occupy remarkably diverse ecological niches²², and it has been reported that they may occur as symbionts of mycorrhizal fungi belonging to the family of Gigasporaceae²³. However, to date the functional significance of the symbiosis has not been elucidated, and previous attempts to cultivate a bacterial symbiont from a fungus or to obtain a symbiont-free fungal host have been unsuccessful²⁴. The association of *Rhizopus* and *Burkholderia* sp. is, to our knowledge, the first example of a bacterial–fungal symbiosis with a clear metabolic function.

In general, pathogenic fungi exert their destructive effects through pathogenicity factors that help the colonizing of living matter, and in all known cases the toxins are biosynthesized by the fungus. In this case, the fungus hosts a bacterial population for the production of the causative agent of rice seedling blight. Rhizoxin, which inhibits mitosis in rice plant cells, efficiently weakens or even kills the plant, and both host and symbiont benefit from nutrients from decaying plant material. It should be noted that this strategy seems not to be restricted to plant infection alone. *R. microsporus* has also been recognized as an occasional cause of human infections, such as cutaneous zygomycoses^{25,26}, and most remarkably, *R. microsporus* CBS 308.87, which was isolated from infected human tissue after a spider bite, also harbours the rhizoxin-producing bacterial symbiont (Table 1, entry 4). In all cases, it is obvious that the fungus profits from the biosynthetic capabilities of the endosymbiont in order to access nutrient sources. Yet the advantage for the bacterial symbiont is not evident, as it may grow, albeit slowly, outside the host and produce the toxin.

This ecologically and medically relevant system raises questions about the acquisition and transmission of the symbiont, as well as the evolution of reciprocal adaptation—in particular, resistance towards rhizoxin. Study of this intriguing tripartite plant–pathogen–symbiont model, as well as the cloning and engineering of rhizoxin biosynthetic genes for drug development, are in progress in our laboratory.

METHODS

Bacterial strains and culture conditions. *R. microsporus* van Tieghem var. *chinensis* (ATCC 62417) and *Rhizopus* sp. F-1360 (ATCC 20577) were obtained from the American Type Culture Collection; *R. microsporus* var. *microsporus* strains CBS 699.68 and CBS 308.87, and *R. microsporus* var. *chinensis* (Saito) Schipper & Stalpers (CBS 631.82) were obtained from the Centraalbureau voor Schimmelcultures; *R. microsporus* van Tieghem var. *chinensis* (DSMZ 1834) was obtained from the Deutsche Sammlung von Mikroorganismen und Zellkulturen. *Rhizopus* strains were grown on potato glucose agar medium at 30 °C. Liquid fermentation was usually carried out in 500 ml Erlenmeyer flasks with 100 ml seed medium (1% corn starch, 0.5% glycerol, 1% gluten meal, 1% dried yeast, 1% corn steep liquor and 1% CaCO₃, pH 6.5) at 30 °C with agitation between 40 and 80 r.p.m. For cultivation of strain CBS 308.87, malt extract peptone medium was used. For generation of symbiont-free *R. microsporus*, the strain was constantly cultivated in the presence of ciprofloxacin (20–40 µg ml⁻¹, Bayer AG).

Isolation of bacterial symbionts. Mycelium from a growing *R. microsporus* culture (20 ml) were transferred to a 50 ml centrifuge tube and mechanically sheared by pipetting. Broken cells were submitted to two rounds of gradient centrifugation (500 r.p.m./32 g for 5 min, 4,000 r.p.m./2,060 g for 20 min). After each centrifugation step, aliquots of the supernatant were plated onto nutrient agar plates (Nutrient Broth Enriched, Serva Electrophoresis; 5.0 g peptone, 3.0 g meat extract, 15.0 g agar, 1,000 ml water) and incubated at 30 °C.

Laser microscopy. Mycelium of a growing culture of *R. microsporus* (0.5 ml) was transferred to an Eppendorf tube containing 0.5 ml 0.85% NaCl. An aliquot of the mixture (10 µl) was placed onto a microscope slide and stained with 0.5 µl SYTO 9 green-fluorescent nucleic acid stain and an equal amount of propidium iodide stain from a Live/Dead BacLight Bacterial Viability Kit (Molecular Probes – L7007) and incubated in the dark for 15 min. Then, ProLong Gold antifade reagent (Molecular Probes – P36935) was added to enhance fluorescence. The sample was incubated for 20 min and analysed using a Zeiss LSM 510 Meta Laser Microscope at 480/500 nm.

DNA isolation, PCR and sequencing. Metagenomic DNA from fungal species was obtained using the method as described¹⁴. Amplification of KS genes from fungal metagenomic DNA was accomplished using the degenerate

primers KSF1 and KSR1 as previously described²⁷. 16S rDNA was amplified using proofreading Taq polymerase Quiagen and universal primers 16S for (CCGAATTCGTCGACAACAGAGTTTGATCCTGGCTCAG) and 16Srev (CCCCGGGATCCAAAGCTTACGGCTACCTTGTACGACTT). PCR conditions were 33 cycles at 96 °C for 1 min, 52 °C for 1 min, and 72 °C for 2 min. PCR products of expected size (about 1.5 kb) were purified using QIAEX II Gel Extraction Kit (QIAGEN) and cloned into pGEM-T easy vector (Promega) for sequencing as described in the manufacturer's instructions. Sequencing was performed on an ABI PRISM 3100 Genetic Analyser with at least threefold coverage.

HPLC-MS analyses. Mycelium of a 4-d-old liquid fermentation (100 ml) was harvested and exhaustively extracted overnight with ethyl acetate (1:1 v/v). The organic phase was filtered, dried over sodium sulphate and concentrated under reduced pressure. The residue was re-dissolved in methanol (500 µl). HPLC analyses were carried out on a Shimadzu Class-VP 6.12 SP5 HPLC equipped with a Nucleosil 100 C18 column (5 µm particle size) using an acetonitrile/0.1% TFA (or formic acid for HPLC-MS) gradient (0.5–100% in 22 min) for elution at a flow rate of 2.0 ml min⁻¹. Rhizoxin and its analogues were monitored using a photodiode array detector, and by ESI-CID-MSⁿ (electrospray ionization in-source collision-induced dissociation mass spectrometry) on a Finnigan LCQ benchtop mass spectrometer equipped with an electrospray ion source and ion-trap mass analyser. The rhizoxin reference was obtained from Sigma.

Received 10 May; accepted 1 July 2005.

1. Furuya, S., Kurata, M. & Saito, T. Studies on *Rhizopus* sp. causing growth injury of young rice seedlings and its chemical control. *Proc. Assoc. Plant Protect. Shikoku* **9**, 49–55 (1974).
2. Ibaragi, T. Studies on rice seedling blight. I. Growth injury caused by *Rhizopus* sp. under high temperature. *Ann. Phytopathol. Soc. Jpn* **39**, 141–144 (1973).
3. Gho, N., Sato, Z., Yaoita, T. & Aoyagi, K. Studies on the control of *Rhizopus* in the nursery cases of rice seedlings. 5. Influence of a phytotoxic substance produced by *Rhizopus* on growth of rice. *Proc. Assoc. Plant Protect. Shikoku* **26**, 90–94 (1978).
4. Noda, T., Hashiba, T. & Sato, Z. The structural changes in young swollen roots of rice seedlings infected with *Rhizopus chinensis* Saito. *Ann. Phytopathol. Soc. Jpn* **46**, 40–45 (1980).
5. Iwasaki, S. *et al.* Studies on macrocyclic lactone antibiotics. VII. Structure of a phytotoxin “rhizoxin” produced by *Rhizopus chinensis*. *J. Antibiot.* **37**, 354–362 (1984).
6. Iwasaki, S. *et al.* Studies on macrocyclic lactone antibiotics. VIII. Absolute structures of rhizoxin and a related compound. *J. Antibiot.* **39**, 424–429 (1986).
7. Koga-Ban, Y., Niki, T., Nagamura, Y., Sasaki, T. & Minobe, Y. cDNA sequences of three kinds of beta-tubulins from rice. *DNA Res.* **2**, 21–26 (1995).
8. Takahashi, M. *et al.* Studies on macrocyclic lactone antibiotics. XI. Anti-mitotic and anti-tubulin activity of new antitumor antibiotics, rhizoxin and its homologues. *J. Antibiot.* **40**, 66–72 (1987).
9. Tsuruo, T. *et al.* Rhizoxin, a macrocyclic lactone antibiotic, as a new antitumor agent against human and murine tumour cells and their vincristine-resistant sublines. *Cancer Res.* **46**, 381–385 (1986).
10. Jordan, A., Hadfield, J. A., Lawrence, N. J. & McGown, A. T. Tubulin as a target for anticancer drugs: agents which interact with the mitotic spindle. *Med. Res. Rev.* **18**, 259–296 (1998).
11. Jennessen, J. *et al.* Secondary metabolite and mycotoxin production by the *Rhizopus microsporus* group. *J. Agric. Food Chem.* **53**, 1833–1840 (2005).
12. Kobayashi, H., Iwasaki, S., Yamada, E. & Okuda, S. Biosynthesis of the antimitotic antitumour antibiotic rhizoxin by *Rhizopus chinensis*; Origins of the carbon atoms. *J. Chem. Soc., Chem. Commun.* 1702–1703 (1986).
13. Hopwood, D. A. Genetic contributions to understanding polyketide synthases. *Chem. Rev.* **97**, 2465–2497 (1997).
14. Nicholson, T. P. *et al.* Design and utility of oligonucleotide gene probes for fungal polyketide synthases. *Chem. Biol.* **8**, 157–178 (2001).
15. Piel, J., Hui, D., Fusetani, N. & Matsunaga, S. Targeting modular polyketide synthases with iteratively acting acyltransferases from metagenomes of uncultured bacterial consortia. *Environ. Microbiol.* **6**, 921–927 (2004).
16. Roberge, M. *et al.* Cell-based screen for antimitotic agents and identification of analogues of rhizoxin, eleutherobin, and paclitaxel in natural extracts. *Cancer Res.* **60**, 5052–5058 (2000).
17. Kenny, D. J., Russell, P., Rogers, D., Eley, S. M. & Titball, R. W. *In vitro* susceptibilities of *Burkholderia mallei* in comparison to those of other pathogenic *Burkholderia* spp. *Antimicrob. Agents Chemother.* **43**, 2773–2775 (1999).
18. Hildebrand, M. *et al.* Approaches to identify, clone, and express symbiont bioactive metabolite genes. *Nat. Prod. Rep.* **21**, 122–124 (2004).
19. Piel, J. Metabolites from symbiotic bacteria. *Nat. Prod. Rep.* **21**, 519–538 (2004).
20. Piel, J. A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of *Paederus* beetles. *Proc. Natl Acad. Sci. USA* **99**, 14002–14007 (2002).

21. Piel, J. *et al.* Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge *Theonella swinhoei*. *Proc. Natl Acad. Sci. USA* **101**, 16222–16227 (2004).
22. Coenye, T. & Vandamme, P. Diversity and significance of *Burkholderia* species occupying diverse ecological niches. *Environ. Microbiol.* **5**, 719–729 (2003).
23. Bianciotto, V. *et al.* An obligately endosymbiotic mycorrhizal fungus itself harbors obligately intracellular bacteria. *Appl. Environ. Microbiol.* **62**, 3005–3010 (1996).
24. Bonfante, P. Plants, mycorrhizal fungi and endobacteria: a dialog among cells and genomes. *Biol. Bull.* **204**, 215–220 (2003).
25. Kobayashi, M. *et al.* Cutaneous zygomycosis: a case report and review of Japanese reports. *Mycoses* **44**, 311–315 (2001).
26. Ribes, J. A., Vanover-Sams, C. L. & Baker, D. J. Zygomycetes in human disease. *Clin. Microbiol. Rev.* **13**, 236–301 (2000).
27. He, J. & Hertweck, C. Iteration as programmed event during polyketide formation; molecular analysis of the aureothin biosynthesis gene cluster. *Chem. Biol.* **10**, 1225–1232 (2003).
28. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. The

ClustalX windows interface: flexible strategies for multiple alignment aided by quality analysis tools. *Nucleic Acids Res.* **24**, 4876–4882 (1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Perner for MS and HPLC-MS measurements, G.-M. Schwinger for strain cultivation, I. Löschmann for sequencing and A. Hartmann for practical help using the laser microscope. We are grateful to J. Piel for advice and discussions, and H. G. Floss and A. Brakhage for critically reading earlier versions of the manuscript. Financial support by the Leibniz Gemeinschaft is acknowledged.

Author Information The 16S rDNA nucleotide sequences of *Rhizopus* sp. symbionts have been deposited at the EMBL Nucleotide Sequence Database under the accession numbers AJ938141–AJ938144. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.H. (Christian.Hertweck@hki-jena.de).

Characterization of the 1918 influenza virus polymerase genes

Jeffery K. Taubenberger¹, Ann H. Reid^{1†}, Raina M. Lourens^{1†}, Ruixue Wang¹, Guozhong Jin¹
& Thomas G. Fanning¹

The influenza A viral heterotrimeric polymerase complex (PA, PB1, PB2) is known to be involved in many aspects of viral replication and to interact with host factors¹, thereby having a role in host specificity^{2,3}. The polymerase protein sequences from the 1918 human influenza virus differ from avian consensus sequences at only a small number of amino acids, consistent with the hypothesis that they were derived from an avian source shortly before the pandemic. However, when compared to avian sequences, the nucleotide sequences of the 1918 polymerase genes have more synonymous differences than expected, suggesting evolutionary distance from known avian strains. Here we present sequence and phylogenetic analyses of the complete genome of the 1918 influenza virus^{4–8}, and propose that the 1918 virus was not a reassortant virus (like those of the 1957 and 1968 pandemics^{9,10}), but more likely an entirely avian-like virus that adapted to humans. These data support prior phylogenetic studies suggesting that the 1918 virus was derived from an avian source¹¹. A total of ten amino acid changes in the polymerase proteins consistently differentiate the 1918 and subsequent human influenza virus sequences from avian virus sequences. Notably, a number of the same changes have been found in recently circulating, highly pathogenic H5N1 viruses that have caused illness and death in humans and are feared to be the precursors of a new influenza pandemic. The sequence changes identified here may be important in the adaptation of influenza viruses to humans.

Influenza A viruses cause annual outbreaks in humans and domestic animals. Periodically, new strains emerge in humans that cause global pandemics. The severe 'Spanish' influenza pandemic of 1918–1919 infected hundreds of millions, and resulted in the death of approximately 50 million people¹². We have previously used phylogenetic analyses to help understand the origin of the pandemic virus^{8,11}; functional studies to understand the pathogenicity of the 1918 virus are underway^{6,13–17}. Recent data have shown that viral constructs bearing the 1918 haemagglutinin gene are pathogenic in a mouse model, but the genetic basis of this observation has not yet been mapped^{6,13–17}. The overall goals of this project have been to understand the origin and unusual virulence of the 1918 influenza virus.

The influenza virus A polymerase functions as a heterotrimer formed by the PB2, PB1 and PA proteins (see ref. 1 for a review). An additional small open reading frame has recently been identified, coding for a peptide (PB1-F2) that is thought to play a role in virus-induced cell death¹⁸. It is not yet clear how the polymerase complex must change to adapt to a new host³. A single amino acid change in PB2, E627K, was shown (1) to be important for mammalian adaptation^{2,3}, (2) to distinguish highly pathogenic avian influenza (HPAI) H5N1 viruses in mice¹⁹, and (3) to be present in the single

fatal human infection during the HPAI H7N7 outbreak in the Netherlands in 2003 (ref. 20), and in some recent H5N1 isolates from humans in Vietnam and Thailand and wild birds in China^{21–23}.

The open reading frame sequences of segment 1 (PB2), segment 2 (PB1) and segment 3 (PA) of A/Brevig Mission/1/1918, and theoretical translations of the four identified reading frames, are shown in Supplementary Fig. 1a–c. The 1918 PB2 protein contained five changes from the avian consensus sequence (Table 1). Of these, A199S is in the area mapped as the PB1 binding site, and the L475M change is in a nuclear localization signal^{24–26}. Three other changes at residues 567, 627 and 702 occur at sites that are not in known functional domains.

The 1918 PB1 protein differed from the avian consensus by seven residues (one of which is shown in Table 1; see also Supplementary Fig. 2). Of these, K54R is in the overlapping binding domains for complementary (c)RNA and viral (v)RNA. Changes at residues 375, 383 and 473 all occur in between the four conserved polymerase motifs in the cRNA binding domain²⁷, and changes at residues 576, 645 and 654 occur in the vRNA binding domain²⁸.

Seven changes were noted in the 1918 PA protein compared with the avian consensus (four of which are shown in Table 1, the other three being C241Y, K312R and I322V). The C241Y change occurs in a nuclear localization signal, but the other six changes (at residues 55, 100, 312, 322, 382 and 552) occur at sites outside of known functional domains^{24–26}.

Representative phylogenetic analyses of the three polymerase genes are shown in Figs 1–3. The 1918 human pandemic viral polymerase genes were compared to representative avian influenza genes with regards to transition/transversion (T_i/T_v) ratio, synonymous/non-synonymous (S/N) ratio, and the numbers of differences at fourfold degenerate sites (defined in ref. 11). T_i/T_v ratios for most comparisons using the 1918 viral genes and representative sequences of either North American or Eurasian avian genes yielded values between 2 and 4. This range was similar to that observed for comparisons of various avian genes with one another, except for the PB1 gene. For PB1, comparisons of the 1918 viral gene with avian virus PB1 genes was always close to 2, whereas comparisons of various avian genes with one another were in the range of 6–10. There were fewer transversions in comparisons between avian PB1 genes than in comparisons between avian PB1 and 1918 human virus PB1, probably reflecting that transversions more often lead to non-synonymous changes.

S/N ratios for most comparisons using the 1918 viral genes and representative sequences of either North American or Eurasian avian genes usually yielded values in the range of 7–16 for both the PA and PB2 genes, as is the case for most avian versus avian PA and PB2 gene comparisons. Like the T_i/T_v ratios, the S/N ratios were somewhat

¹Department of Molecular Pathology, Armed Forces Institute of Pathology, Rockville, Maryland 20850, USA. [†]Present addresses: Board on Life Sciences, The National Academies, 6th Floor, 500 Fifth Street N.W., Washington DC 20001, USA (A.H.R.); University of Iowa, Roy J. and Lucille A. Carver College of Medicine, 200 CMAB, Iowa City, Iowa 52242, USA (R.M.L.).

Table 1 | Amino acid residues distinguishing human and avian influenza polymerases

Gene	Residue no.	Avian	1918	Human H1N1	Human H2N2	Human H3N2	Classical swine	Equine
PB2	199	A	S	S	S	S	S	A
PB2	475	L	M	M	M	M	M	L
PB2	567	D	N	N	N	N	D	D
PB2	627	E	K	K	K	K	K	E
PB2	702	K	R	R*	R	R	R	K
PB1	375	N/S/T†	S	S	S	S	S	S
PA	55	D	N	N	N	N	N	N
PA	100	V	A	A	A	A	V	A
PA	382	E	D	D	D	D	D	E
PA	552	T	S	S	S	S	S	T

* All human H1N1 PB2 sequences have an Arg residue at position 702, except that two out of three A/PR/8/34 sequences have a Lys residue.

† The majority of avian sequences have an Asn residue at position 375 of PB1, 18% have a Ser residue, 13% a Thr residue.

higher with the PB1 gene (most of the comparisons yielded ratios in the range of 16–25), owing to a smaller number of non-synonymous changes in comparisons of avian PB1 genes with one another. These findings may reflect a more conservative evolution of PB1 in birds.

A subset of synonymous differences occurs at sites that are fourfold degenerate (that is, where a substitution with any base does not result in an amino acid replacement). As these sites are not subject to selective pressure at the protein level, base substitutions at many fourfold degenerate sites may accumulate rapidly. If influenza virus genes have been evolving in birds for long enough to reach evolutionary stasis, as is suggested by the high S/N ratios described above, one would predict that at many of the sites where fourfold degeneracy is possible, all four bases would be present in the avian clade unless the constraints of RNA secondary structure limit the accumulation of synonymous changes. In fact, when avian sequences from geographically distinct lineages (North American versus European) were compared, the per cent difference at fourfold degenerate sites yielded values in the 27–38% range. In contrast, calculating the per cent difference at fourfold degenerate sites in comparisons of the 1918 viral PA, PB1 and PB2 gene sequences with avian sequences yielded consistently higher values (range 41–51%) for all three genes. As with the other 1918 genes¹¹, this suggests that the donor source of the 1918 virus was in evolutionary isolation from those avian influenza viruses currently represented in the databases.

Emphasizing the avian-like nature of the 1918 influenza virus polymerase proteins, out of 19 total amino acid changes from the avian consensus, there are only 10 amino acid positions (out of 2,232 total codons) that consistently distinguish the 1918 and subsequent human polymerase proteins PB2, PB1 and PA from their avian influenza counterparts (these are defined as changes from avian sequences in the 1918 virus that are maintained without change in subsequent human viruses) (Table 1). It is likely that these changes have an important role in human adaptation. Seven of these ten changes were previously noted in an alignment between avian and human influenza polymerases³. What follows is a comparison between the 1918 virus changes and recent H5N1 isolates, in order to evaluate possible examples of parallel evolution in the adaptation of avian influenza viruses to humans.

In the PB2 protein, five changes distinguish the human isolates from avian sequences (Table 1). Out of 253 available PB2 sequences from human H1N1, H2N2 and H3N2 isolates, these five changes are almost completely preserved, with the exception that two recent H3N2 isolates have the avian Lys residue at position 702. Only a small number of avian influenza isolates show any of these five changes, and it is intriguing that almost all of these isolates are from HPAI H5N1 or H7N7 viruses, or from the H9N2 lineage that infected a small number of humans in China in the late 1990s (ref. 29). Only 5 out of 282 available avian PB2 sequences have a Ser residue at position 199, four of these being 1997 H5N1 isolates from Hong Kong. The A199S change was also found in 5 out of 18 H5N1 strains isolated from humans (all five were from the 1997 Hong Kong outbreak). Of the avian viruses, 36 out of 336 have an Arg residue

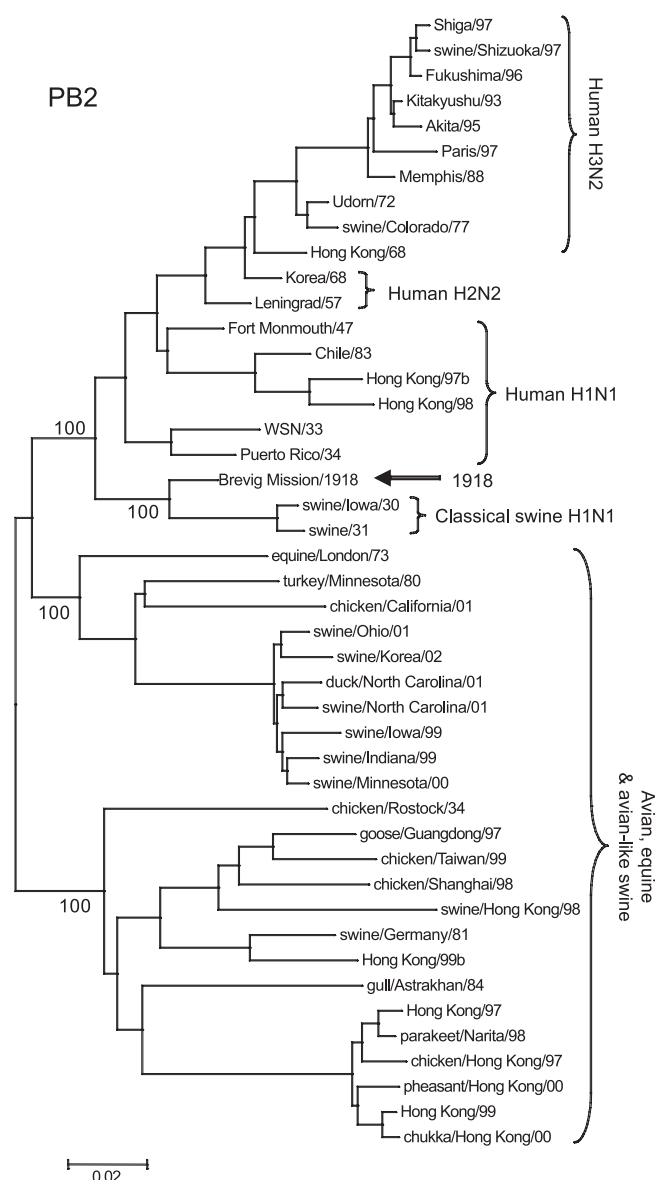


Figure 1 | Phylogenetic tree of the PB2 gene. Sequences were aligned and analysed for phylogenetic relationships using the NJ algorithm, with the proportion of sequence differences as the distance measure. Bootstrap values (100 replications) for key nodes are shown (for clarity, identical and nearly identical sequences have been removed from the trees). Major clades are identified with large brackets. The arrow identifies the position of the 1918 PB2 gene sequence. A distance bar is shown below the tree. Influenza strain abbreviations used in the analyses are listed in Supplementary Table 1.

at position 702, 30 of which are H9N2 isolates from China around 1996–2000, and 5 are H5N1 isolates from Hong Kong in 1997 and 2001. Out of 18 available 1997 H5N1 strains isolated from humans, three have the K702R change.

Perhaps most interestingly, the 1918 virus and subsequent human isolates have a Lys residue at position 627. This residue has been implicated in host adaptation^{2,3}, and has previously been shown to be crucial for high pathogenicity in mice infected with the 1997 H5N1 virus¹⁹. Of the avian isolates, 19 out of 345 have a Lys residue at position 627, 18 of which are HPAI H5N1 or H7N7 avian influenza viruses. Sixteen of these were recently characterized H5N1 isolates from a die-off of wild waterfowl around Qinghai Lake in western China in 2005 (ref. 21). In human H5N1 isolates, 11 out of 37 have the E627K change: A/Hong Kong/483/1997 and A/Hong Kong/485/1997, four out of six isolates from Vietnam in 2004 (ref. 22), and two out of three isolates from Thailand in 2004 (ref. 23). The E627K mutation was seen in six out of seven H5N1 isolates from Thai tigers in 2004, and was also present in the H7N7 virus responsible for the single human fatality during the HPAI H7N7 outbreak in the Netherlands in 2003 (ref. 20). It was not noted in the contemporaneous chicken isolates.

At position 475, only one out of 355 avian isolates has a Met residue (an H5N1 HPAI virus from 2004). Similarly, only one out of 345 avian viruses has an Asn residue at position 567. None of the

human H5N1 isolates has the L475M or the D567N changes. None of the available H5N1 or H7N7 sequences has more than one of the proposed human-adaptive PB2 changes determined for the 1918 virus.

The PA protein shows a similar pattern: four residues consistently differ between 1918 and subsequent human isolates and the avian consensus sequence (Table 1). Three other changes (C241Y, K312R and I322V) distinguish 1918, H1N1 and H2N2 human isolates, but most H3N2 isolates have the avian amino acid at these positions. Of 295 available sequences from human H1N1, H2N2 and H3N2 isolates, all have Asn at position 55 (except A/WSN/33), Ala at position 100 and Ser at position 552. Only 5 out of 295 human isolates have the avian Glu residue at position 382. Notably, these five isolates make up a minor clade of recent H3N2 isolates that have a number of unusual changes from typical human H3N2 viruses³⁰. When avian influenza sequences are analysed, none (out of 209 sequences) has Asn at position 55 or Ser at position 552. Only 8 out of 209 avian PA protein sequences show the V100A change: six recent H6N2 isolates from chickens in California, and two HPAI 2002 H5N1 duck isolates from China. Of the 209 avian sequences, five have an Asp residue at position 382, including two HPAI H5N2 isolates from chickens in Mexico in 1994.

The PB1 gene segment was replaced by reassortment in both the

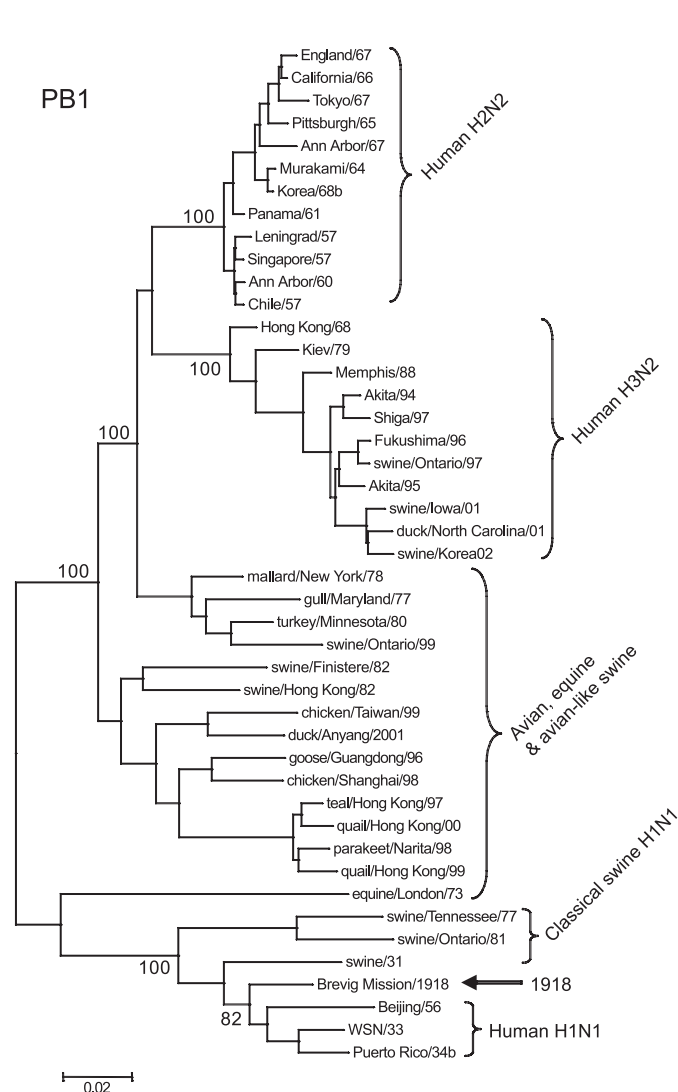


Figure 2 | Phylogenetic tree of the PB1 gene. Sequences were aligned and analysed as detailed in the legend to Fig. 1.

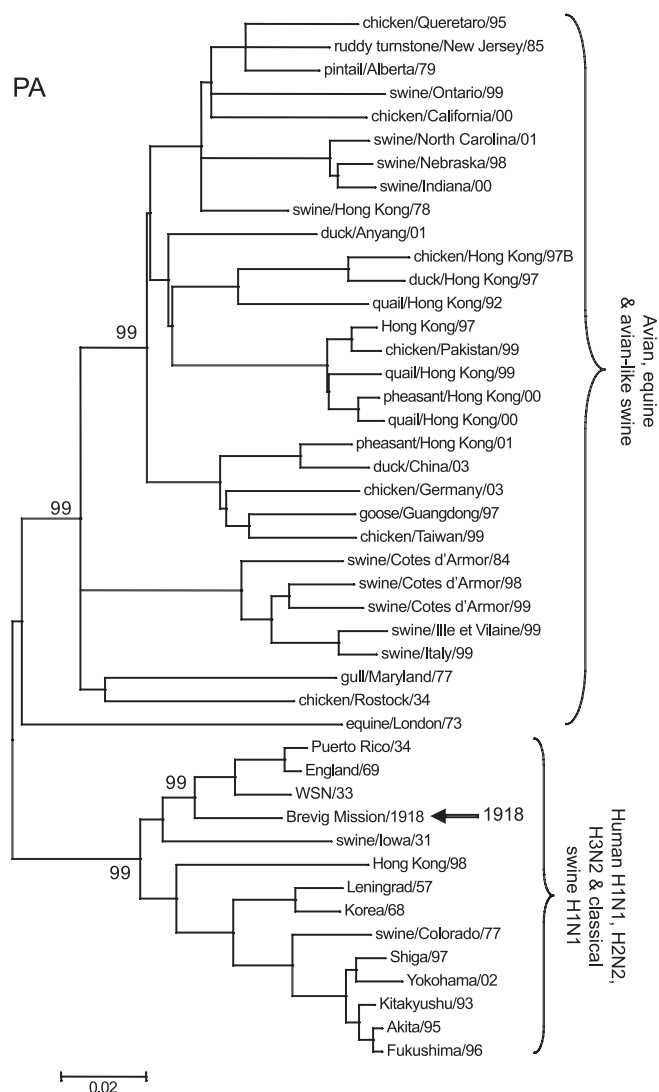


Figure 3 | Phylogenetic tree of the PA gene. Sequences were aligned and analysed as detailed in the legend to Fig. 1.

1957 and 1968 pandemics⁹. We compared the PB1 protein from the 1918 human virus with those of the avian-derived PB1 segments from the 1957 and 1968 pandemics. Human H1N1, H2N2 and H3N2 viruses derived from the 1918, 1957 and 1968 pandemics, respectively, each possessed a uniquely derived avian-like PB1 gene segment, and so we sought to identify any parallel changes that might shed light on human adaptation. The three human pandemic PB1 proteins differ from the avian consensus by only 4–7 residues each (Supplementary Fig. 2). Only one of these changes is shared among the pandemic isolates: an N375S change. This change to a serine residue is also found in swine and equine influenza A isolates. With few exceptions, all human influenza PB1 proteins have Ser at this site. Of 230 human influenza sequences, only two H1N1 isolates (A/FM/47 and A/Beijing/1956) and the ‘minor clade’ H3N2 isolates described above have the avian Asn residue³⁰. In contrast, although this residue is maintained in almost all mammalian isolates, it is variable among avian PB1 proteins. Of 293 avian isolates, 66% have the consensus Asn residue at position 375, 18% have a Ser residue and 12% have a Thr residue.

The data presented here highlight the marked conservation of the PB1 protein in avian influenza viruses. PB1 functions as an RNA-dependent RNA polymerase, and so it is reasonable to hypothesize that its enzymatic function is optimal in this conserved form. In humans, the PB1 proteins experience linear change over time. Indeed, PB1 in humans acquires ~0.4 amino acid changes per year. As there is such strong antigenic selection on human viruses, it is possible that although the observed changes in PB1 are selectively beneficial with respect to antigenicity, they are mildly deleterious to enzyme function. Such complex fitness trade-offs are thought to be commonplace in RNA virus evolution. Supporting this hypothesis, a recent study examining combinations of avian and human influenza polymerases showed that the most efficient influenza transcriptional activity *in vitro* was seen with an avian-derived PB1, even if the PB2, PA and NP proteins were from a human virus³. Acquiring an avian PB1 by reassortment might provide a replicative advantage to the new virus, possibly explaining why both of the last two pandemics and the 1918 influenza virus all had very avian-like PB1 proteins.

Both the 1957 and 1968 pandemic influenza viruses were avian/human reassortants in which 2–3 avian gene segments were reassorted with the then-circulating, human-adapted virus^{9,10}. Unlike the 1957 and 1968 pandemics, however, the 1918 virus was most likely not a human/avian reassortant virus, but rather an avian-like virus that adapted to humans *in toto*^{8,11}. On the basis of amino acid replacement rates in human influenza virus polymerase genes, it is possible that these segments were circulating in human influenza viruses as early as 1900. However, proof that the 1918 virus did not retain gene segments from the previously circulating human influenza A strain would require discovery of a sample of the pre-1918 virus from archival material. The donor source, although avian-like at the protein level, may have come from a subset of avian influenza viruses not currently represented in the sequence databases and may have been in evolutionary isolation.

The fact that amino acid changes identified in the 1918 analysis are also seen in HPAI strains of H5N1 and H7N7 avian viruses that have caused fatalities in humans is intriguing, and suggests that these changes may facilitate virus replication in human cells and increase pathogenicity. It is possible that the high pathogenicity of the 1918 virus was related to its emergence as a human-adapted avian influenza virus. These changes may reflect a process of parallel evolution as avian influenza A viruses mutate in response to adaptational pressures, and suggest that the genetic basis of avian influenza virus adaptation to humans can be mapped.

METHODS

RNA isolation, amplification and sequencing. RNA was isolated from frozen 1918 human lung tissue using Trizol (Invitrogen) according to the

manufacturer's instructions. Each fragment was reverse transcribed, amplified, and sequenced at least twice. Reverse transcription polymerase chain reaction (RT-PCR), isolation of products and sequencing have been previously described⁴. Lists of primers and primer sequences are available upon request. Replicate RT-PCR reactions from independently produced RNA preparations gave identical sequence results. The 2,280-nucleotide complete coding sequence of PB2 was amplified in 33 overlapping fragments. The 2,274-nucleotide coding sequence of PB1 was amplified in 33 overlapping fragments. The 2,151-nucleotide coding sequence of PA was amplified in 32 overlapping fragments. The PCR products ranged in size from 77–138 bp.

Phylogenetic analyses. Phylogenetic analyses of the three polymerase genes were done using standard methods. We generated trees using the neighbour-joining (NJ) algorithm, with proportion of differences as the distance measure using MEGA version 2.1. Character evolution was analysed with the MacClade program after a parsimony analysis using PAUP version 4.0 beta, using ACTRAN as the optimization method. Trees were also generated using maximum-likelihood with midpoint rooting. All algorithms generated comparable trees, with major clades representing human, classical swine and avian-like viruses (NJ trees shown in Figs 1–3; complete data set available upon request). Polymerase segment sequences used in this analysis were obtained from GenBank and the Influenza Sequence Databank (ISD). (See Supplementary Table 1 for a list of sequences used.) For the PB2 gene, 83 sequences were used, all of which were full length. For the PB1 gene, 91 sequences were used, three of which were not full length. For the PA gene, 105 sequences were used, six of which were not full length.

Received 30 June; accepted 19 September 2005.

1. Fodor, E. & Brownlee, G. G. in *Influenza* (ed. Potter, C. W.) 1–29 (Elsevier, Amsterdam, 2002).
2. Subbarao, E. K., London, W. & Murphy, B. R. A single amino acid in the PB2 gene of influenza A virus is a determinant of host range. *J. Virol.* **67**, 1761–1764 (1993).
3. Naffakh, N., Massin, P., Escriviou, N., Crescenzo-Chaigne, B. & van der Werf, S. Genetic analysis of the compatibility between polymerase proteins from human and avian strains of influenza A viruses. *J. Gen. Virol.* **81**, 1283–1291 (2000).
4. Reid, A. H., Fanning, T. G., Hultin, J. V. & Taubenberger, J. K. Origin and evolution of the 1918 “Spanish” influenza virus hemagglutinin gene. *Proc. Natl Acad. Sci. USA* **96**, 1651–1656 (1999).
5. Reid, A. H., Fanning, T. G., Janczewski, T. A. & Taubenberger, J. K. Characterization of the 1918 “Spanish” influenza virus neuraminidase gene. *Proc. Natl Acad. Sci. USA* **97**, 6785–6790 (2000).
6. Basler, C. F. et al. Sequence of the 1918 pandemic influenza virus nonstructural gene (NS) segment and characterization of recombinant viruses bearing the 1918 NS genes. *Proc. Natl Acad. Sci. USA* **98**, 2746–2751 (2001).
7. Reid, A. H., Fanning, T. G., Janczewski, T. A., McCall, S. & Taubenberger, J. K. Characterization of the 1918 “Spanish” influenza virus matrix gene segment. *J. Virol.* **76**, 10717–10723 (2002).
8. Reid, A. H., Fanning, T. G., Janczewski, T. A., Lourens, R. & Taubenberger, J. K. Novel origin of the 1918 pandemic influenza virus nucleoprotein gene segment. *J. Virol.* **78**, 12462–12470 (2004).
9. Kawaoka, Y., Krauss, S. & Webster, R. G. Avian-to-human transmission of the PB1 gene of influenza A viruses in the 1957 and 1968 pandemics. *J. Virol.* **63**, 4603–4608 (1989).
10. Scholtissek, C., Rohde, W., Von Hoyningen, V. & Rott, R. On the origin of the human influenza virus subtypes H2N2 and H3N2. *Virology* **87**, 13–20 (1978).
11. Reid, A. H., Taubenberger, J. K. & Fanning, T. G. Evidence of an absence: the genetic origins of the 1918 pandemic influenza virus. *Nature Rev. Microbiol.* **2**, 909–914 (2004).
12. Johnson, N. P. & Mueller, J. Updating the accounts: global mortality of the 1918–1920 “Spanish” influenza pandemic. *Bull. Hist. Med.* **76**, 105–115 (2002).
13. Geiss, G. K. et al. Cellular transcriptional profiling in influenza A virus-infected lung epithelial cells: the role of the nonstructural NS1 protein in the evasion of the host innate defense and its potential contribution to pandemic influenza. *Proc. Natl Acad. Sci. USA* **99**, 10736–10741 (2002).
14. Tumpey, T. M. et al. Existing antivirals are effective against influenza viruses with genes from the 1918 pandemic virus. *Proc. Natl Acad. Sci. USA* **99**, 13849–13854 (2002).
15. Tumpey, T. M. et al. Pathogenicity and immunogenicity of influenza viruses with genes from the 1918 pandemic virus. *Proc. Natl Acad. Sci. USA* **101**, 3166–3171 (2004).
16. Kash, J. C. et al. The global host immune response: contribution of HA and NA genes from the 1918 Spanish influenza to viral pathogenesis. *J. Virol.* **78**, 9499–9511 (2004).
17. Kobasa, D. et al. Enhanced virulence of influenza A viruses with the haemagglutinin of the 1918 pandemic virus. *Nature* **431**, 703–707 (2004).
18. Chen, W. et al. A novel influenza A virus mitochondrial protein that induces cell death. *Nature Med.* **7**, 1306–1312 (2001).

19. Shinya, K. *et al.* PB2 amino acid at position 627 affects replicative efficiency, but not cell tropism, of Hong Kong H5N1 influenza A viruses in mice. *Virology* **320**, 258–266 (2004).
20. Fouchier, R. A. *et al.* Avian influenza A virus (H7N7) associated with human conjunctivitis and a fatal case of acute respiratory distress syndrome. *Proc. Natl Acad. Sci. USA* **101**, 1356–1361 (2004).
21. Chen, H. *et al.* Avian flu: H5N1 virus outbreak in migratory waterfowl. *Nature* **436**, 191–192 (2005).
22. Li, K. S. *et al.* Genesis of a highly pathogenic and potentially pandemic H5N1 influenza virus in eastern Asia. *Nature* **430**, 209–213 (2004).
23. Puthavathana, P. *et al.* Molecular characterization of the complete genome of human influenza H5N1 virus isolates from Thailand. *J. Gen. Virol.* **86**, 423–433 (2005).
24. Toyoda, T., Adyshev, D. M., Kobayashi, M., Iwata, A. & Ishihama, A. Molecular assembly of the influenza virus RNA polymerase: determination of the subunit–subunit contact sites. *J. Gen. Virol.* **77**, 2149–2157 (1996).
25. Masunaga, K., Mizumoto, K., Kato, H., Ishihama, A. & Toyoda, T. Molecular mapping of influenza virus RNA polymerase by site-specific antibodies. *Virology* **256**, 130–141 (1999).
26. Ohtsu, Y., Honda, Y., Sakata, Y., Kato, H. & Toyoda, T. Fine mapping of the subunit binding sites of influenza virus RNA polymerase. *Microbiol. Immunol.* **46**, 167–175 (2002).
27. Biswas, S. K. & Nayak, D. P. Mutational analysis of the conserved motifs of influenza A virus polymerase basic protein 1. *J. Virol.* **68**, 1819–1826 (1994).
28. Gonzalez, S. & Ortin, J. Distinct regions of influenza virus PB1 polymerase subunit recognize vRNA and cRNA templates. *EMBO J.* **18**, 3767–3775 (1999).
29. Guo, Y. J. *et al.* Characterization of the pathogenicity of members of the newly established H9N2 influenza virus lineages in Asia. *Virology* **267**, 279–288 (2000).
30. Holmes, E. C. *et al.* Whole-genome analysis of human influenza A virus reveals multiple persistent lineages and reassortment among recent H3N2 viruses. *PLoS Biol.* **3**, e300 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The research described in this report was done using stringent safety precautions to protect the laboratory workers, the environment and the public from this virus. The intention of this research is to provide the basis for understanding how influenza pandemic strains form and to help ascertain the risk of future influenza pandemics. This study was partially supported by a grant to J.K.T. from the National Institutes of Health, and by intramural funds from the Armed Forces Institute of Pathology. The opinions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the US Department of the Army or the US Department of Defense.

Author Contributions J.K.T. planned the project, and A.H.R., R.M.L., R.W. and G.J. generated the sequence data. J.K.T., A.H.R. and T.G.F. performed data analysis. J.K.T. wrote the manuscript.

Author Information Coding sequences of the PB2, PB1 and PA genes have been deposited in GenBank under accession numbers DQ208309, DQ208310 and DQ208311, respectively. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.K.T. (taubenberger@afip.osd.mil).

LETTERS

***In vivo* analysis of quiescent adult neural stem cells responding to Sonic hedgehog**

Sohyun Ahn^{1†} & Alexandra L. Joyner^{1,2}

Sonic hedgehog (Shh) has been implicated in the ongoing neurogenesis in postnatal rodent brains^{1,2}. Here we adopted an *in vivo* genetic fate-mapping strategy, using Gli1 (GLI-Kruppel family member) as a sensitive readout of Shh activity, to systematically mark and follow the fate of Shh-responding cells in the adult mouse forebrain. We show that initially, only a small population of cells (including both quiescent neural stem cells and transit-amplifying cells) responds to Shh in regions undergoing neurogenesis. This population subsequently expands markedly to continuously provide new neurons in the forebrain. Our study of the behaviour of quiescent neural stem cells provides *in vivo* evidence that they can self-renew for over a year and generate multiple cell types. Furthermore, we show that the neural stem cell niches in the subventricular zone and dentate gyrus are established sequentially and not until late embryonic stages.

Two regions of sustained neurogenesis in the adult mammalian nervous system are the subventricular zone (SVZ) of the lateral ventricles, which provides interneurons of the olfactory bulb (OB) via the rostral migratory stream (RMS), and the subgranular zone (SGZ) of the dentate gyrus (DG) within the hippocampus, which generates granule neurons (reviewed in refs 3–6). Although previous studies have shown that the secreted protein Sonic hedgehog (Shh) increases cell proliferation in the SVZ and SGZ^{1,2} and is required for normal proliferation in the SVZ², it has not been determined whether Shh acts on putative quiescent neural stem cells and/or fast-dividing transit-amplifying cells.

In order to investigate whether neural stem cells normally respond to Shh in the adult mouse brain, we used genetic techniques to identify cells that respond to Shh. First, the SVZ and SGZ of adult mice ($n = 3$) were analysed for co-expression of Gli1-*nlacZ*, which is a sensitive readout of positive Shh signalling^{7,8}, and glial fibrillary acidic protein (GFAP), which marks neural stem cells^{9,10} as well as mature astrocytes. Shh-responding (nuclear lacZ⁺) cells in the ventral SVZ and SGZ co-expressed GFAP (25.3% and 18.8%, respectively), suggesting that neural stem cells respond to Shh activity (Fig. 1b, c). In addition, 18% of Shh-responding cells in the SVZ and 36% in the SGZ co-expressed polysialic acid-neural cell adhesion molecule (PSA-NCAM, Supplementary Table 1), a neuroblast/neural precursor marker. Co-labelling with *Dlx2* (distal-less homeobox 2; ref. 11, data not shown) or *Olig2* (oligodendrocyte transcription factor 2; ref. 12, Supplementary Fig. 1; 57% of Shh-responding cells), which are markers of transit-amplifying cells, showed that fast-dividing transit-amplifying cells also respond to Shh signalling. However, no Shh-responding cells were found in the RMS and OB, the target structures of the cells generated in the SVZ, suggesting that only neural stem cells and their immediate precursor cells respond to Shh activity.

The only definitive test to show that Shh signals to neural stem cells *in vivo* would be to mark the Shh-responding cells and determine whether they self-renew and continuously give rise to mature neurons. To mark and follow Shh-responding cells in the SVZ and SGZ *in vivo*, we performed genetic fate mapping using *Gli1-CreER^{T2}* mice, in which an inducible Cre recombinase (CreER^{T2}) is expressed from *Gli1* (ref. 13). Upon tamoxifen treatment of *Gli1-CreER^{T2};R26R* mice, CreER^{T2} is transiently activated for ~30 h, resulting in permanent expression of cytoplasmic lacZ from the *R26R* locus in *Gli1*-expressing cells and all their progeny^{13,14}. Initially, a small number of Shh-responding cells in the SVZ and SGZ of adult *Gli1-CreER^{T2};R26R* mice ($n = 3$) were labelled (Fig. 1e, h). Owing to the mosaic nature of our genetic fate mapping approach, not all Shh-responding cells are labelled following tamoxifen treatment^{13,14}. However, marker expression analysis of the initial population revealed that the distribution of different marked cell types was similar to that of *Gli1-nlacZ* mice (Supplementary Table 1). Thus, the initially marked population faithfully represents the Shh-responding cells in the SVZ and SGZ.

The long-term fate of the marked cells analysed six months after tamoxifen treatment revealed a marked increase in the number of labelled cells in *Gli1-CreER^{T2};R26R* mice ($n = 5$) in both the SVZ and DG (Fig. 1g, i and data not shown). In addition, although no labelled cells were detected in the RMS or OB at the initial time point (Fig. 1d and data not shown), long-term fate mapping showed labelled cells throughout the RMS and OB (Fig. 1f and data not shown), demonstrating that the small initial population of Shh-responding cells in the SVZ was amplified to generate labelled cells in the OB that had migrated via the RMS. Also, in the DG of the hippocampus, many labelled cells had integrated into deeper layers of the granular zone after six months (Fig. 1i, black bracket) compared with the initial Shh-responding population that was primarily confined to the SGZ (Fig. 1h, red bracket). Taken together, our long-term fate-mapping results indicate that Shh-responding cells in the SVZ and SGZ include neural stem cells that can generate progeny over time in their respective target structures.

To determine whether Shh-responding cells include the quiescent neural stem cells that rarely divide, rather than only the fast-dividing transit-amplifying cells with a limited lifespan, we used AraC, an anti-mitotic reagent that effectively kills fast-dividing cells in the SVZ and SGZ while sparing quiescent neural stem cells¹⁵. Two days after tamoxifen treatment, AraC was infused for one week into the brain of *Gli1-CreER^{T2};R26R* mice using a mini-osmotic pump. At the end of AraC treatment (day 0), PSA-NCAM immunoreactivity in whole-mounts of the lateral wall of the lateral ventricles ($n = 3$) showed that the neuroblasts were almost completely depleted (Fig. 2b and data

¹Howard Hughes Medical Institute, Developmental Genetics Program, Skirball Institute of Biomolecular Medicine and Department of Cell Biology, and ²Department of Physiology and Neuroscience, New York University School of Medicine, 540 First Avenue, New York, New York 10016, USA. †Present address: Unit on Developmental Neurogenetics, Laboratory of Mammalian Genes and Development, National Institute of Child Health and Human Development, National Institutes of Health, 9000 Rockville Pike, Building 6B, Bethesda, Maryland 20892, USA.

not shown), confirming the efficacy of AraC treatment. Elimination of fast-dividing precursors/neuroblasts by AraC treatment was confirmed on sections through the SVZ and DG, where only one of 69 and one of 63 lacZ⁺ cells (respectively) showed PSA-NCAM immunoreactivity (Supplementary Table 1). In addition, *Dlx2* expression was abolished in the SVZ and was markedly diminished in the RMS (Supplementary Fig. 2 and data not shown).

At the end of AraC treatment, $92.2 \pm 1.0\%$ and $93.7 \pm 4.7\%$ of the remaining labelled cells in the SVZ and SGZ, respectively, co-expressed GFAP (data not shown), indicating that quiescent neural stem cells respond to Shh signalling *in vivo*. Strikingly, within one week of AraC removal ($n = 5$), the number of labelled cells in the SVZ and DG of AraC-treated mice had increased more than in

control mice ($n = 4$) (Fig. 2h–j and data not shown). No labelled cells were observed in the OB at the end of AraC treatment (Fig. 2c), but labelled cells were present in the granule cell layer of the OB one week after AraC removal (Fig. 2g). The extent of labelling in AraC-treated mice, while continuing to expand, was comparable with control mice after two months ($n = 3$ each, data not shown), three months ($n = 4$ each, data not shown) and six months ($n = 3$ each, Fig. 2k–n). These results provide additional strong evidence that Shh-responding cells in the SVZ and SGZ that survive AraC treatment are rarely dividing neural stem cells that respond to this insult by increasing cell proliferation and the production of transit-amplifying cells.

Even though neural stem cells isolated from the SVZ and SGZ have characteristics of stem cells *in vitro*, including self-renewal and generation of multiple cell types^{10,16,17}, it has not been demonstrated whether quiescent neural stem cells *in vivo* self-renew or generate cell types other than neurons in their respective target structures^{18,19}. Previous lineage-tracing experiments have relied on the rare infection of neural stem cells and transit-amplifying cells with replication-defective retroviruses^{20,21} or BrdU labelling of progenitor

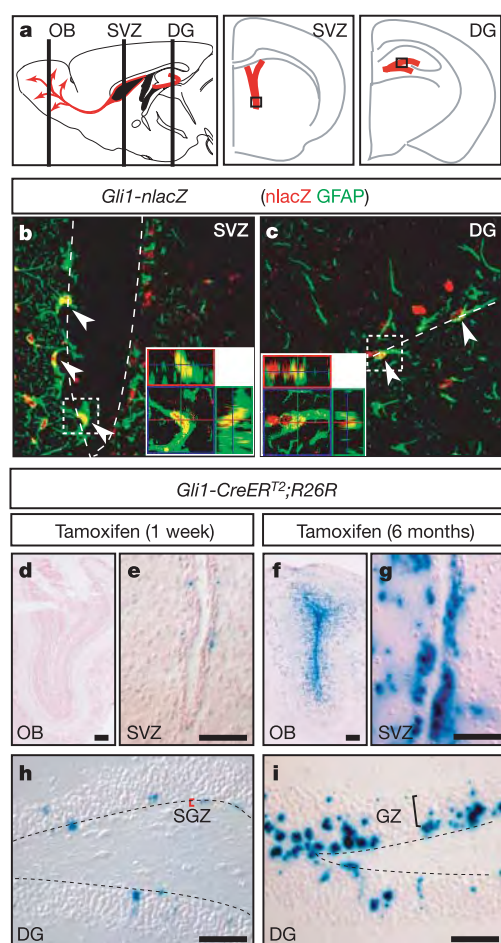


Figure 1 | Neural stem cells respond to Shh and expand over time *in vivo*.

a, Left panel shows a schematic sagittal section of adult mouse brain with the neurogenic regions and migratory paths indicated in red lines. Black area indicates the lateral ventricle. The levels of coronal sections shown in panels **b–i** are indicated by vertical lines. The two right panels show schematics of hemi-coronal sections of adult forebrain at the level of the subventricular zone (SVZ) and the dentate gyrus (DG). **b, c**, Confocal images of single optical slices show expression of GFAP (green; neural stem cell) and nlacZ (red; Shh-responding cell) in the SVZ (**b**) and DG (**c**) of two-month-old *Gli1-nlacZ* mice. Arrowheads indicate double-labelled cells and the insets show the orthogonal analysis of a representative cell marked in a box. **d–i**, Genetic fate-mapping of Shh-responding cells over time using *Gli1-CreERT2;R26R* mice. Short-term fate mapping (1 week) shows no labelled cells in the olfactory bulb (OB, **d**) and few cells in the SVZ (**e**) and DG (**h**). Six months after treatment with tamoxifen, many labelled cells are found in the OB (**f**), SVZ (**g**) and DG (**i**). The red bracket in **h** indicates the subgranular zone (SGZ) and the black bracket in **i** indicates the granular zone (GZ). Scale bars, 20 μm (**b, c**), 50 μm (**e–i**).

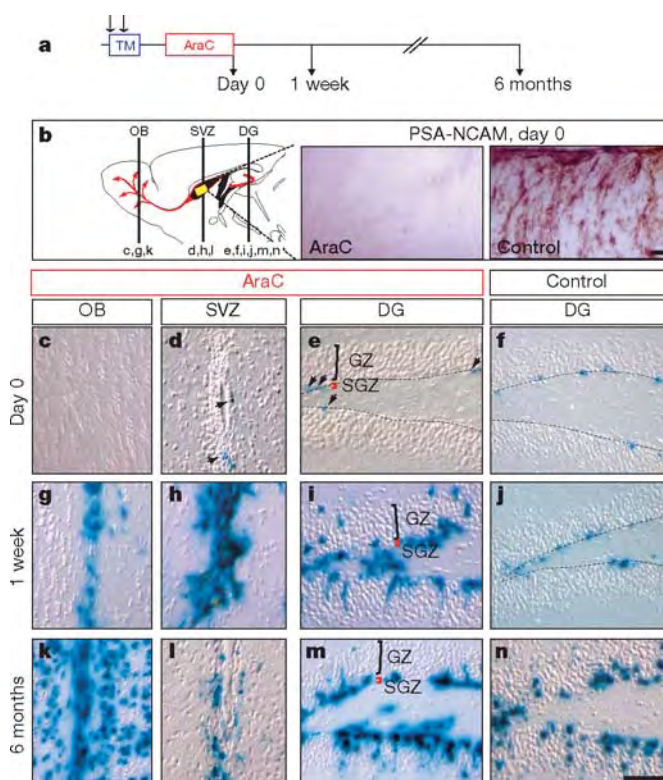


Figure 2 | Quiescent neural stem cells respond to Shh signalling.

a, A schematic of the experimental time course. Tamoxifen (TM) was administered on two consecutive days (arrows) and AraC or saline was infused for one week. Forebrains of operated mice were analysed at the indicated survival time. **b**, Left panel shows a schematic sagittal section of adult mouse forebrain. Vertical lines indicate the level of coronal sections shown in panels **c–n**. Yellow box indicates the area of the lateral wall of the SVZ analysed for PSA-NCAM (neuroblast marker) immunoreactivity shown in the right panels. One week of AraC treatment effectively abolished the proliferating neuroblasts from the lateral wall of the lateral ventricle compared with control treatment. **c–n**, Fate-mapped cells were analysed at the end of AraC treatment (day 0, **c–e**), and one week (**g–i**) and six months (**k–m**) after AraC treatment, and compared to saline-treated cells at the corresponding time points (**f, j, n**). Shh-responding cells survived the anti-mitotic treatment with AraC and expanded over time in the SVZ and DG. Arrows indicate the few cells labelled after AraC treatment. Red brackets indicate the subgranular zone (SGZ) and black brackets indicate the granular zone (GZ) of the DG. Scale bars, 50 μm .

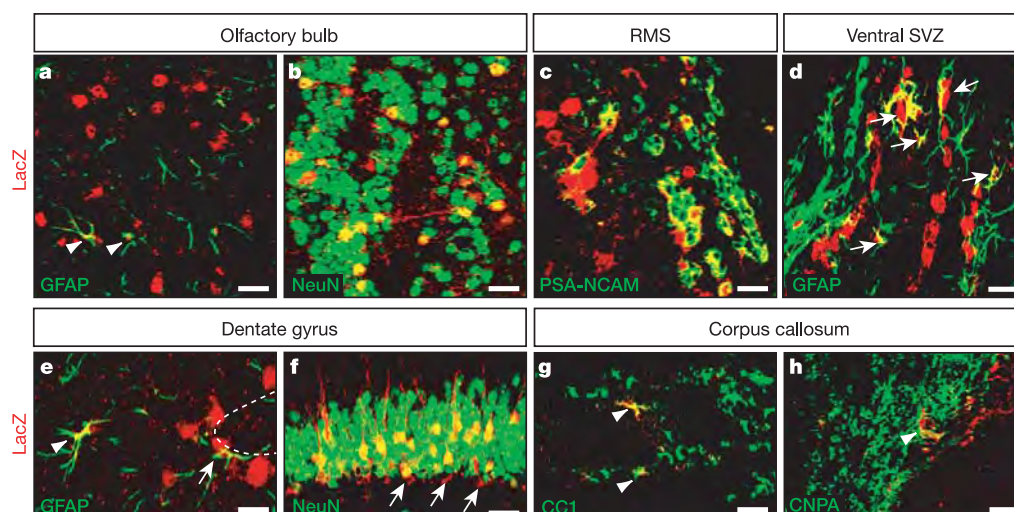


Figure 3 | Shh-responding neural stem cells generate multiple cell types *in vivo*. Long-term fate-mapped cells (one year after tamoxifen treatment) were analysed for co-expression of lacZ (red) and the indicated molecular markers (green). **a–f**, The vast majority of labelled cells are interneurons (NeuN^+) in the olfactory bulb (**b**) and the dentate gyrus (**f**), as well as a few astrocytes (GFAP^+ , arrowheads in **a**, **e**). Fate-mapped neuroblasts

(PSA-NCAM^+) are found in the rostral migratory stream (RMS, **c**). Quiescent neural stem cells (GFAP^+) are found in the SVZ (**d**) and SGZ (**e**, arrows). **g**, **h**, Few fate-mapped cells in the corpus callosum are oligodendrocytes (arrowheads). CC1 and CNPA are oligodendrocyte markers. Scale bars, 10 μm (**a–e**, **g**, **h**), 20 μm (**f**).

cells in their last division^{22,23} to determine the long-term fate of neural stem cells. We took advantage of our efficient method for marking quiescent neural stem cells in the adult to determine their long-term fate after one year (Fig. 3). We first concentrated on the self-renewal capacity of the quiescent neural stem cells that respond to Shh signalling. The percentage of labelled cells co-expressing GFAP in the SVZ had increased significantly ($P < 0.05$) one year after marking (from $23.2 \pm 1.6\%$ initially to $34.2 \pm 4.7\%$ after one year, values mean $\pm$ s.d.) and in the SGZ (from $18.8 \pm 3.3\%$ initially to $30.3 \pm 5.3\%$ after one year). In addition, the absolute number of labelled cells co-expressing GFAP in the SVZ and SGZ had increased, suggesting that Shh-responding neural stem cells had self-renewed to generate more neural stem cells in both AraC-treated and control mice (Fig. 3d, e and Supplementary Table 1). To further demonstrate the ability of Shh-responding neural stem cells to self-renew, *Gli1-CreER^{T2};R26R* mice ($n = 2$) that were treated with AraC following tamoxifen administration were allowed to recover for three months and then subjected to another round of AraC treatment. Even after a second round of eliminating the fast-dividing precursors, the remaining neural stem cells that were marked before the first AraC treatment survived and generated more PSA-NCAM^+ neuroblasts (Supplementary Fig. 3). Together, these data show that quiescent neural stem cells respond to Shh signalling and self-renew *in vivo*.

In addition to the GFAP^+ cells that remain in the SVZ, many PSA-NCAM and lacZ co-expressing cells were observed in the dorsolateral SVZ (data not shown), where neuroblasts exit the SVZ to enter the RMS en route to the OB. In the anterior RMS, most of the lacZ^+ cells were co-labelled with PSA-NCAM (Fig. 3c). Thus, quiescent neural stem cells in the SVZ actively produce new neuroblasts even one year after marking. In the OB, the vast majority of labelled cells were neurons as shown by their co-expression of lacZ and NeuN ($92.3 \pm 4.5\%$, Fig. 3b) as well as their elaborate dendritic arborizations. In the periglomerular region, there was extensive co-expression of lacZ and γ -aminobutyric acid (GABA) or tyrosine hydroxylase, whereas labelled granular neurons typically expressed only GABA (Supplementary Fig. 4). We also detected a few GFAP -expressing fate-mapped cells within the OB ($3.5 \pm 0.5\%$, $n = 3$) (Fig. 3a). This suggests that although most of the progeny generated from the SVZ become interneurons of the OB, some also become

astrocytes in the OB. In addition, there were a few marked oligodendrocytes in the corpus callosum (Fig. 3g, h). In the DG, the majority of fate-mapped cells co-expressed NeuN ($82.2 \pm 7.8\%$), indicating that they had become granule neurons in the deeper layers of the granular zone (Fig. 3f). A few lacZ -expressing cells in the deeper layer of the granular zone also expressed GFAP ($3.3 \pm 0.7\%$), suggesting that the neural stem cells had also generated astrocytes *in vivo* (Fig. 3e and Supplementary Table 1). Our results therefore demonstrate that quiescent neural stem cells have the potential to generate multiple cell types *in vivo*.

Our genetic fate-mapping technique provides an opportunity to explore the genetic history of neural stem cells and the formation of neural stem cell niches. To determine whether adult neural stem cells in the SVZ and DG are derived from cells that had previously responded to Shh in the embryo, we analysed the long-term fate of Shh-responding cells marked at various embryonic stages. No

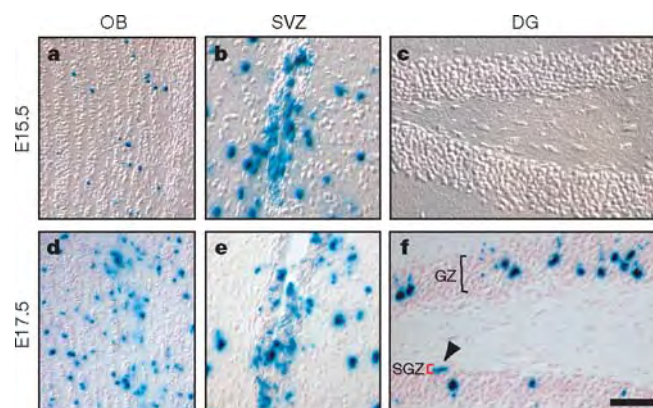


Figure 4 | Shh-responding neural stem cells are established during late embryogenesis in mice. Tamoxifen was administered on E15.5 and E17.5 to mark cells responding to Shh at E16–17 and E18–19, respectively. One-month-old mouse forebrains show labelled cells in the OB (**a**, **d**) and ventral SVZ (**b**, **e**) when marked at both time points. Labelled subgranular zone (SGZ; red bracket and arrowhead) cells of the DG were seen by marking at E18–19 (**f**) but not at E16–17 (**c**). Scale bar, 50 μm . GZ, granular zone.

labelled cells were found in the adult SVZ and SGZ when tamoxifen was administered between embryonic days (E)7.5 and E11.5 (data not shown). The appearance of marked cells was first observed in the SVZ when tamoxifen was given at E15.5, and labelled cells in the SGZ did not appear until tamoxifen was given at E17.5 (Fig. 4). This shows that Shh does not act on forebrain neural stem cells until late embryogenesis, and further suggests that the two stem cell niches in the forebrain are established sequentially in mice. In addition, we found that the Shh-responding cells in the DG marked at E17.5 showed a different behaviour from cells marked in the adult. Cells marked as a consequence of tamoxifen treatment on E17.5 populated deeper layers of the granular zone of the DG than cells marked in the adult (compare Figs 1i, 2n and 4f, black brackets), suggesting a temporal restriction in the extent of integration of newly generated cells as the brain matures.

Our genetic fate-mapping studies demonstrate that Shh first acts on stem cell niches in the SVZ and SGZ at late embryonic stages. Quiescent neural stem cells are then set aside and are regulated by Shh signalling. We present *in vivo* evidence that the neural stem cells can self-renew for at least a year and generate multiple cell types over time. Hedgehog signalling has recently been implicated in the stem cell biology associated with tissue repair and the progression of tumours in many non-neural tissues^{24–26}. Our *in vivo* genetic fate-mapping approach provides a unique opportunity to determine whether in fact stem cells regulated by Hedgehog signalling have a role in cell replenishment of various organs, and to elucidate the mechanism by which they contribute to tissue repair and cancer.

METHODS

Fate mapping and drug treatment. Two- to three-month-old *Gli1-CreER^{T2}/+*;R26R/R26R adult mice were administered 10 mg tamoxifen (Sigma) in corn oil by oral gavage on two consecutive days using feeding needles (Fine Science Tools). For AraC treatment (two days after tamoxifen treatment), a mini-osmotic pump (model 1007D, Durect) was implanted to infuse either saline or 2% AraC (Sigma) for one week as described¹⁵. Brains of mice killed at the end of AraC or control treatment were designated as day 0 samples. For long-term survival experiments, the pump was removed one week after implantation and mice were later killed at the indicated survival time. The second application of AraC was performed in the same way as the first AraC treatment.

Immunostaining and analysis of brain tissues. Mice were perfused with 4% paraformaldehyde (PFA), processed for frozen sections, and stained with X-gal as described on <http://saturn.med.nyu.edu/research/dg/joynerlab/>. For immunofluorescent staining, dissected brains were fixed in 4% PFA overnight at 4°C and 50-µm vibratome (Leica) sections were obtained. Free-floating sections were stained overnight at 4°C with goat anti-β-galactosidase (1:500, Biogenesis) and one of the following antibodies: mouse anti-GFAP (1:500), anti-PSA-NCAM (1:300), anti-NeuN (1:200, Chemicon), anti-CC1 (1:20, Calbiochem), anti-CNPase (1:150), anti-MBP (1:200, SMI), rabbit anti-GABA (1:500, Sigma) and anti-Olig2 (1:20,000; ref. 27). Secondary antibodies for double labelling were donkey anti-species conjugated with Alexa 488, Alexa 555 (1:500, Molecular Probes) or FITC (1:200, Jackson ImmunoResearch). Fluorescent images were captured in 1.5-µm optical sections using a confocal laser-scanning microscope (LSM 510, Zeiss) and processed using Adobe Photoshop. Orthogonal analysis was performed to confirm co-expression of two markers. Cells co-expressing lacZ and one of the various markers were counted and divided by total number of lacZ-expressing cells in the indicated region. At least three sections each through the SVZ and DG were analysed from one mouse. Data from at least three mice were pooled together to determine the average and standard deviation (s.d.). Student's *t*-tests were performed to calculate *P* values and to determine whether the results between two time points were significantly different.

Received 4 April; accepted 29 June 2005.

- Lai, K., Kaspar, B. K., Gage, F. H. & Schaffer, D. V. Sonic hedgehog regulates adult neural progenitor proliferation *in vitro* and *in vivo*. *Nature Neurosci.* **6**, 21–27 (2003).
- Machold, R. *et al.* Sonic hedgehog is required for progenitor cell maintenance in telencephalic stem cell niches. *Neuron* **39**, 937–950 (2003).
- Doetsch, F. The glial identity of neural stem cells. *Nature Neurosci.* **6**, 1127–1134 (2003).
- Alvarez-Buylla, A. & Garcia-Verdugo, J. M. Neurogenesis in adult subventricular zone. *J. Neurosci.* **22**, 629–634 (2002).
- Gage, F. H. Mammalian neural stem cells. *Science* **287**, 1433–1438 (2000).
- Alvarez-Buylla, A. & Lim, D. A. For the long run: maintaining germinal niches in the adult brain. *Neuron* **41**, 683–686 (2004).
- Bai, C. B., Auerbach, W., Lee, J. S., Stephen, D. & Joyner, A. L. *Gli2*, but not *Gli1*, is required for initial Shh signalling and ectopic activation of the Shh pathway. *Development* **129**, 4753–4761 (2002).
- Bai, C. B., Stephen, D. & Joyner, A. L. All mouse ventral spinal cord patterning by Hedgehog is Gli dependent and involves an activator function of Gli3. *Dev. Cell* **6**, 103–115 (2004).
- Doetsch, F., Caille, I., Lim, D. A., Garcia-Verdugo, J. M. & Alvarez-Buylla, A. Subventricular zone astrocytes are neural stem cells in the adult mammalian brain. *Cell* **97**, 703–716 (1999).
- Seri, B., Garcia-Verdugo, J. M., McEwen, B. S. & Alvarez-Buylla, A. Astrocytes give rise to new neurons in the adult mammalian hippocampus. *J. Neurosci.* **21**, 7153–7160 (2001).
- Doetsch, F., Petreanu, L., Caille, I., Garcia-Verdugo, J. M. & Alvarez-Buylla, A. EGF converts transit-amplifying neurogenic precursors in the adult brain into multipotent stem cells. *Neuron* **36**, 1021–1034 (2002).
- Hack, M. A., Sugimori, M., Lundberg, C., Nakafuku, M. & Gotz, M. Regionalization and fate specification in neurospheres: the role of Olig2 and Pax6. *Mol. Cell. Neurosci.* **25**, 664–678 (2004).
- Ahn, S. & Joyner, A. L. Dynamic changes in the response of cells to positive hedgehog signalling during mouse limb patterning. *Cell* **118**, 505–516 (2004).
- Zervas, M., Millet, S., Ahn, S. & Joyner, A. L. Cell behaviors and genetic lineages of the mesencephalon and rhombomere 1. *Neuron* **43**, 345–357 (2004).
- Doetsch, F., Garcia-Verdugo, J. M. & Alvarez-Buylla, A. Regeneration of a germinal layer in the adult mammalian brain. *Proc. Natl Acad. Sci. USA* **96**, 11619–11624 (1999).
- Palmer, T. D., Takahashi, J. & Gage, F. H. The adult rat hippocampus contains primordial neural stem cells. *Mol. Cell. Neurosci.* **8**, 389–404 (1997).
- Gage, F. H., Kempermann, G., Palmer, T. D., Peterson, D. A. & Ray, J. Multipotent progenitor cells in the adult dentate gyrus. *J. Neurobiol.* **36**, 249–266 (1998).
- Suhonen, J. O., Peterson, D. A., Ray, J. & Gage, F. H. Differentiation of adult hippocampus-derived progenitors into olfactory neurons *in vivo*. *Nature* **383**, 624–627 (1996).
- Herrera, D. G., Garcia-Verdugo, J. M. & Alvarez-Buylla, A. Adult-derived neural precursors transplanted into multiple regions in the adult brain. *Ann. Neurol.* **46**, 867–877 (1999).
- Yoon, S. O. *et al.* Adenovirus-mediated gene delivery into neuronal precursors of the adult mouse brain. *Proc. Natl Acad. Sci. USA* **93**, 11974–11979 (1996).
- Doetsch, F., Garcia-Verdugo, J. M. & Alvarez-Buylla, A. Cellular composition and three-dimensional organization of the subventricular germinal zone in the adult mammalian brain. *J. Neurosci.* **17**, 5046–5061 (1997).
- Kuhn, H. G., Dickinson-Anson, H. & Gage, F. H. Neurogenesis in the dentate gyrus of the adult rat: age-related decrease of neuronal progenitor proliferation. *J. Neurosci.* **16**, 2027–2033 (1996).
- Kempermann, G., Gast, D., Kronenberg, G., Yamaguchi, M. & Gage, F. H. Early determination and long-term persistence of adult-generated new neurons in the hippocampus of mice. *Development* **130**, 391–399 (2003).
- Berman, D. M. *et al.* Widespread requirement for Hedgehog ligand stimulation in growth of digestive tract tumours. *Nature* **425**, 846–851 (2003).
- Karhadkar, S. S. *et al.* Hedgehog signalling in prostate regeneration, neoplasia and metastasis. *Nature* **431**, 707–712 (2004).
- Watkins, D. N. *et al.* Hedgehog signalling within airway epithelial progenitors and in small-cell lung cancer. *Nature* **422**, 313–317 (2003).
- Arnett, H. A. *et al.* bHLH transcription factor Olig1 is required to repair demyelinated lesions in the CNS. *Science* **306**, 2111–2115 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to thank J. Wolf, K. Harper, D. Stephens, O. Aristizabal and R. Turnbull for technical assistance; M. Fuccillo for experimental advice; P. Soriano for R26R mice; D. Rowitch and C. Stiles for the anti-Olig2 antibody; G. Fishell for his insightful input throughout the project, and M. Zervas and S. Blaess for critical reading of the manuscript. S.A. is supported by a NIH National Research Service Award postdoctoral fellowship and A.L.J. is a Howard Hughes Medical Institute investigator.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.L.J. (joyner@saturn.med.nyu.edu).

LETTERS

Sex-specific peptides from exocrine glands stimulate mouse vomeronasal sensory neurons

Hiroko Kimoto¹, Sachiko Haga¹, Koji Sato¹ & Kazushige Touhara¹

In mammals, social and reproductive behaviours are modulated by pheromones, which are chemical signals that convey information about sex and strain^{1,2}. The vomeronasal organ, located at the base of the nasal septum, is responsible for mediating pheromone information in mice^{3–9}. Two classes of putative pheromone receptor gene families, V1R and V2R, are expressed by vomeronasal sensory neurons in mutually segregated epithelial zones of the vomeronasal organ^{10–14}. Although numerous studies have suggested that pheromones originate from urine^{15–18}, direct recordings of behaving mice have shown that neuronal firing in the vomeronasal system is modulated by physical contact with the facial area¹⁹. Here we identify a male-specific 7-kDa peptide secreted from the extraorbital lacrimal gland. This peptide, which we named exocrine gland-secreting peptide 1 (ESP1), is encoded by a gene from a previously unrecognized large family clustered in proximity to the class I major histocompatibility complex (MHC) region. ESP1 is secreted from the eyes and is transferred to the female vomeronasal organ, where it stimulates V2R-expressing vomeronasal sensory neurons and elicits an electrical response. Our results indicate that mice respond to sex-specific peptides released from exocrine glands through the vomeronasal system during direct contact.

Induction of c-Fos expression in the vomeronasal system can be used as a marker for pheromonal signal input^{9,20–23}. Thus, to investigate pheromonal communication in behaving mice at the molecular level, we examined the induction of c-Fos as a marker for activated vomeronasal sensory neurons (VSNs) after stimulation with soiled bedding. Exposing BALB/c female mice to bedding soiled by adult male conspecifics resulted in a transient induction of c-Fos expression in some VSNs (Fig. 1a and Supplementary Fig. S1). The induction of c-Fos peaked 1.5–3 h after stimulation and returned to basal levels after 12 h, suggesting that continuously stimulated VSNs lose the ability to express c-Fos. Almost all c-Fos-positive neurons colocalize with $G\alpha_o$ -positive neurons in the basal zone of the vomeronasal neuroepithelium (Supplementary Fig. S1b), indicating that responses to soiled bedding were mainly mediated by V2R receptors. Exposure to male-soiled bedding induced a robust expression of c-Fos in female V2R neurons (8.3 ± 1.0 c-Fos-positive cells in $G\alpha_o$ -positive zone, mean $\pm$ s.e.m., $n = 10$), whereas c-Fos induction was not observed in the male vomeronasal organ (VNO) (Fig. 1b and Supplementary Fig. S1b, c). Expression of another immediate early gene, *egr1* (early growth response 1, ref. 22), was also induced in female VSNs in response to male-soiled bedding, consistent with the observations of c-Fos induction (data not shown). Stimulation with female-soiled bedding caused only a small increase in the number of c-Fos-positive neurons in the male VNO (1.2 ± 0.8 , $n = 4$; Fig. 1b). However, free interaction between female and male mice resulted in c-Fos expression in VSNs of both female and male mice (3.6 ± 1.2 in females ($n = 3$) and 5.5 ± 2.6 in

males ($n = 4$), Fig. 1c). Interaction between mice of the same gender did not induce c-Fos expression (Fig. 1c).

Bedding soiled by prepubertal male mice caused only a small increase in the number of c-Fos-positive neurons in the female VNO (0.9 ± 0.5 , $n = 3$), suggesting that a male-specific cue was released by adult mice (Fig. 1d). Exposure of BALB/c females to bedding soiled by adult C57BL/6 and ICR males caused c-Fos expression in VSNs (6.9 ± 1.0 ($n = 6$) and 6.7 ± 0.4 ($n = 3$),

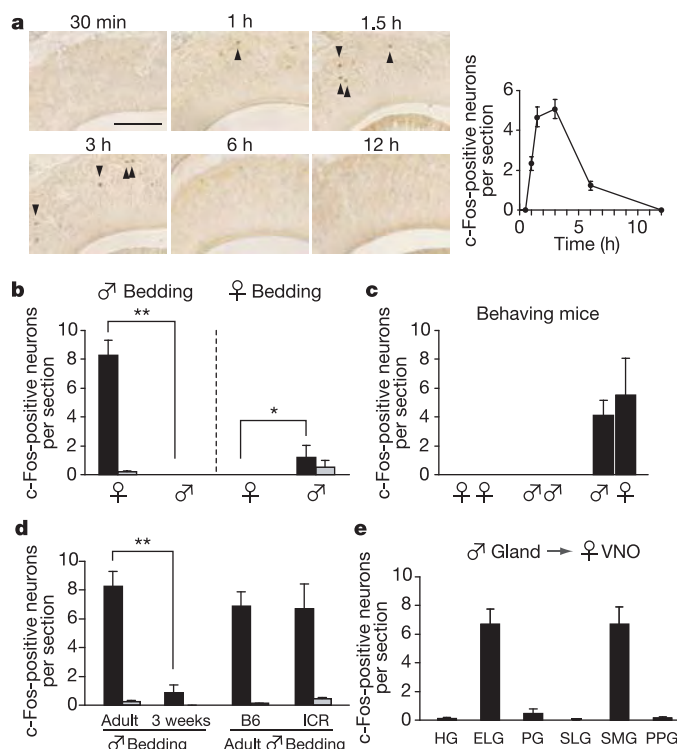


Figure 1 | Induction of c-Fos in VSNs by various stimulants. **a**, Time course of c-Fos induction in female VSNs by male-soiled bedding. Scale bar, 100 μ m. **b**, Stimulation of BALB/c female or male mice with male- or female-soiled bedding ($n = 4$ –10 mice). **c**, c-Fos induction in behaving male and female mice ($n = 3$ –4 mice). **d**, Stimulation of BALB/c females with BALB/c adult or prepubertal (3 weeks of age) mouse male-soiled bedding, or C57BL/6 (B6) or ICR adult male-soiled bedding ($n = 3$ –6 mice). **e**, Screening of glands for secretion of active compound(s). Each stimulation was performed with extract equivalent to one male gland ($n = 3$ –8 mice). The number of c-Fos-positive neurons was counted in 25–27 (**a–d**) or 4–16 (**e**) sections for each mouse. Black bars, $G\alpha_o$ -positive neurons; grey bars, $G\alpha_i$ -positive neurons. All results are shown as mean $\pm$ s.e.m. Data in **b** and **d** were analysed by Mann-Whitney *U*-test. Asterisk, $P < 0.05$; two asterisks, $P < 0.01$.

¹Department of Integrated Biosciences, Graduate School of Frontier Sciences, The University of Tokyo, Chiba 277-8562, Japan.

respectively), indicating that c-Fos induction is not a strain-specific event (Fig. 1d). Urine or putative urine-derived pheromones such as 2-heptanone, 3,4-dehydro-*exo*-brevicomine (DHB) and 2-sec-butyl-4,5-dihydrothiazole (SBT), induced little or no expression of c-Fos (Supplementary Fig. S2). In contrast, shaved fur from male mice contained an activity that induced c-Fos expression in female VSNs (3.6 ± 0.6 , $n = 6$; Supplementary Fig. S2). Furthermore, without direct contact, exposure to soiled bedding did not induce c-Fos (Supplementary Fig. S2). Together, these observations suggest that the male-specific compounds causing c-Fos expression in the female VNO are nonvolatile and originate in a gland of adult male mice rather than from urine.

To explore the source of c-Fos-inducing cues, we examined various exocrine glands that produce substances affecting the VNO, including the Harderian gland (HG), the extraorbital lacrimal gland (ELG), the parotid gland (PG), the sublingual gland (SLG), the submaxillary gland (SMG) and the preputial gland (PPG). The extract from male ELG and SMG induced robust expression of c-Fos in female VSNs (6.7 ± 1.0 ($n = 8$) and 6.7 ± 1.2 ($n = 5$), respectively; Fig. 1e). These results suggest that non-volatile male-specific cues are secreted from the ELG and SMG, resulting in the induction of c-Fos expression in female VSNs. Extracts from six different glands in females caused a slight increase in c-Fos-positive cells in the male VNO (data not shown).

A single V2R neuron is thought to express only one type of V2R receptor^{12,13}. To identify which V2R receptor(s) is expressed in c-Fos-induced VSNs, we designed 12 *in situ* hybridization probes to recognize a total of 36 annotated V2R genes. Of the c-Fos-positive neurons activated by male-soiled bedding, 87% were stained with the V2Rp probe, but not with other probes, including those for V2R1,

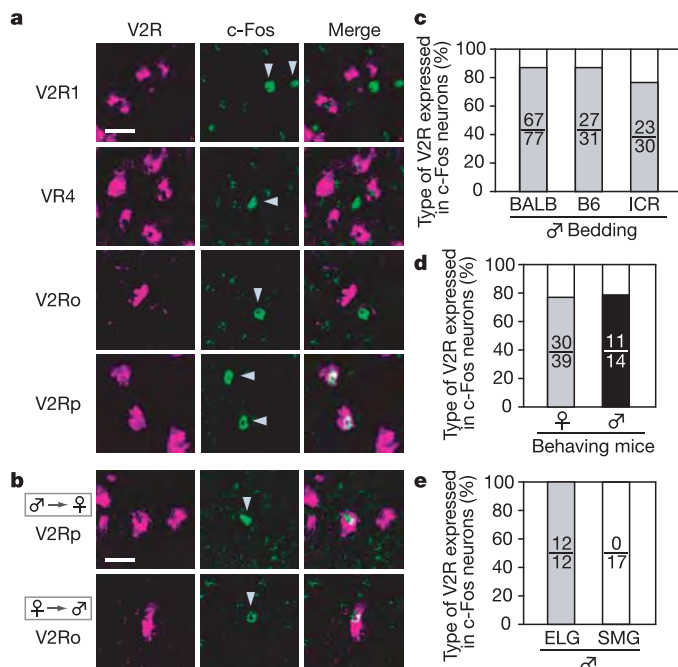


Figure 2 | Identification of V2R receptor genes expressed in c-Fos-positive VSNs. **a, b,** Two-colour fluorescent visualization of V2R receptors (magenta) and c-Fos (green). The female VNO exposed to male-soiled bedding (**a**) or a male mouse (**b**, top row). The male VNO exposed to a female mouse (**b**, bottom row). Scale bars, 20 μ m. **c–e,** The population of V2Rp (grey) and V2Ro (black) neurons out of all c-Fos-positive cells in BALB/c females exposed to BALB/c, C57BL/6 (B6) or ICR male-soiled bedding (**c**), in behaving BALB/c females and males (**d**), and in the female VNO exposed to an extract of male ELG or SMG (**e**). The number of V2Rp or V2Ro neurons and the total number of c-Fos-positive neurons are indicated in each bar (3–12 sections).

VR4 and V2Ro (Fig. 2a, c). Approximately 10% of V2Rp-expressing neurons were c-Fos-positive. The c-Fos-induced cells of female VSNs generated as a result of physical interaction with male mice were also V2Rp-positive (77%) (Fig. 2b, d). In contrast, c-Fos-induced neurons in male mice in response to female mice did not overlap with the V2Rp probe, but 79% did overlap with V2Ro (Fig. 2b, d). These results indicate that male- and female-specific cues are distinct and stimulate different sets of V2Rs. Exposure to C57BL/6 and ICR male-soiled bedding also led to c-Fos induction in female V2Rp-positive neurons (Fig. 2c). All c-Fos-positive neurons generated in response to male ELG extract expressed V2Rp, whereas the cells activated by the SMG extract were distinct from the V2Rp-positive neurons (Fig. 2e).

We next purified and identified male-specific active compounds in the ELG extract. A homogenate of ELG from adult males was loaded onto an anion-exchange DEAE column (Fig. 3a). The flow-through fraction was then subjected to a two-step purification using a reverse-phase C4 column, resulting in the resolution of two c-Fos-inducing peaks (Fig. 3a). Amino-terminal peptide sequence analysis, mass spectrometry and a genome search revealed that there was an open reading frame on chromosome 17 corresponding to an identical N-terminal sequence and to the molecular masses of the three

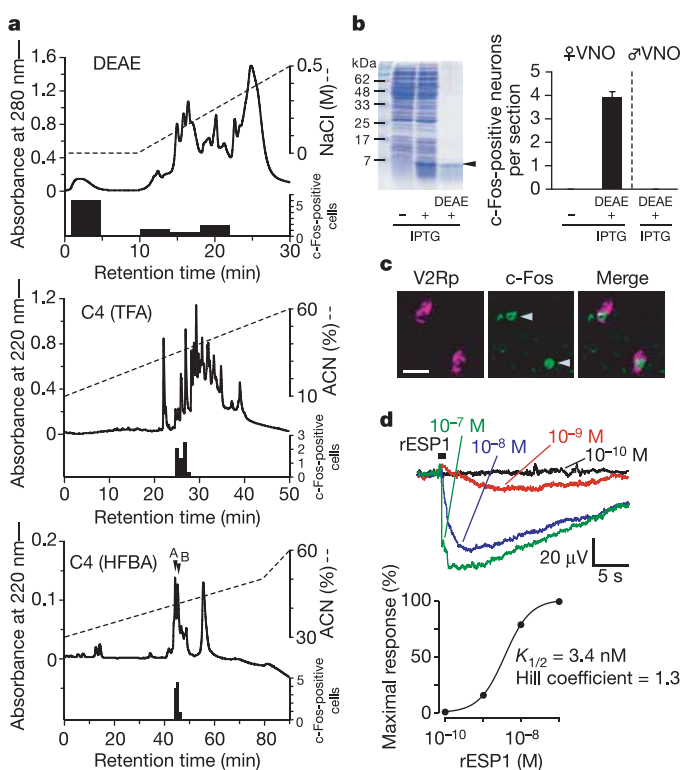


Figure 3 | Purification and identification of c-Fos-inducing peptides.

a, Three-step purification of c-Fos-inducing activity using a DEAE column and a C4 column in the presence of TFA and HFBA. The elution profiles were obtained from samples equivalent to 5 (top panel), 48 (middle) and 43 (bottom) ELGs. Fractions A and B were analysed by peptide sequencing and mass spectrometry (see Supplementary Fig. S3a, b). **b,** c-Fos-inducing activity of rESP1 purified to homogeneity by the DEAE column, in the presence or absence of isopropyl thiogalactoside (IPTG). Data are mean $\pm$ s.e.m.; $n = 8$ mice and 14–16 sections per mouse. **c,** Two-colour fluorescent visualization. All c-Fos-positive neurons (green) stimulated by rESP1 express V2Rp (magenta) in female mice (42 out of 42 neurons, 6 sections). Scale bar, 20 μ m. **d,** Dose-dependent electrical responses of the female VNO to rESP1. A dose-response plot of peak magnitude was obtained from the experiment shown in the upper panel, and a smooth curve was fitted using the Hill equation. Data are representative of three recordings from three females.

peptides (Supplementary Fig. S3a, b). We named this new peptide ESP1, for exocrine gland-secreting peptide.

A DEAE-purified recombinant ESP1 (rESP1; from Ala 1 to Gly 64 in Supplementary Fig. S3b) induced c-Fos expression in the female VNO (3.8 ± 0.4 , $n = 8$) but not in the male VNO (Fig. 3b). c-Fos-induction by rESP1 was observed in a dose-dependent manner (Supplementary Fig. S3c). c-Fos-positive neurons activated by rESP1 overlapped with V2Rp-positive neurons (Fig. 3c), demonstrating that ESP1 was the desired active substance. Furthermore, rESP1 evoked a negative field potential in the female VNO (Supplementary Fig. S3d), suggesting that the ESP1 signal was converted to an electronic event in the VNO. The electrical responses elicited by rESP1 were dose-dependent, with a threshold concentration of 10^{-10} – 10^{-9} M (Fig. 3d). The male VNO also showed an electrical response to ESP1 (data not shown), suggesting that although male mice do not respond to their own ESP1 in the c-Fos assay, the male VNO does respond electrically to ESP1. This result is consistent with

the fact that V2Rp receptors are expressed in both male and female mice. The most likely reason for this lack of c-Fos induction is that male VSNs are continuously stimulated by the presence of their own ESP1, and have therefore lost the ability to express c-Fos (consistent with the time course results in Fig. 1a). However, we cannot exclude the possibility of involvement of other V2Rs in the electrical response. Although the difference between the signal transduction mechanisms leading to c-Fos induction and the electrical response remains to be determined, our results suggest that the male-specific peptide ESP1 elicits an electrical response in the vomeronasal pathway.

The expression of the ESP1 gene in the ELG was male-specific and was observed after four weeks of age (Fig. 4a). No other exocrine gland or tissue expressed the ESP1 gene. ESP1 seems to be secreted to the outside after signal-sequence cleavage and N- and C-terminal deletion by proteolysis in the ELG (Fig. 4b and Supplementary Fig. S3b). We found an additional 23 genes encoding homologous peptides, suggesting that ESP1 is a member of a new multigene family (Fig. 4b). Ten of these genes (ESP2–11) seem to be expressed in the ELG, the HG and/or the SMG of both male and female mice (Fig. 4c). It is an intriguing possibility that these homologous peptides might electrically stimulate VSNs expressing other V2Rs. Interestingly, the ESP gene family is clustered within a 3-Mb region of the telomeric end of MHC class I loci (Fig. 4b); it should be noted that the M10 family of MHC class I molecules is co-expressed with the V2Rs^{8,24}.

ESP1 was detected in male tear fluid by western blot analysis using an anti-ESP1 antibody (Fig. 4d). Thus, we suggest that the male ELG secretes the male-specific peptide ESP1 from the eyes, and the peptide is then transferred to the female VNO through physical contact upon investigation of the facial areas (Fig. 4e). In aquatic-breeding newts and frogs, sex-specific peptides 10–25 amino acids in length are known to be used for sexual communication^{25,26}, whereas in a terrestrial salamander the male delivers 10–22 kDa pheromones to the female nares during courtship^{27,28}. Each animal seems to have developed an evolutionarily unique strategy for transmitting non-volatile sex-specific cues.

Identification of the sex-specific peptide ESP1, which appears to be a candidate sex pheromone in mice and a member of a previously unrecognized multigene family, opens a new avenue for understanding sexual communication and gender discrimination. The male signal ESP1, and an as-yet-unidentified female signal, are received by different types of V2R neurons in the VNO. This leads to activation of distinct neuronal pathways, which may result in construction of information about sex in the brain. Our findings shed light on the evolution of mechanisms underlying pheromonal communication and its relevance to determinants of individual identity, and provide insight into the neuroendocrinology of defining sex differences in mammals.

METHODS

Preparation of pheromone source and stimulation of mice. Mice (SLC) were housed under a 12 h light/12 h dark cycle (light on at 8:00). All experiments were performed with 10-week-old sexually naive BALB/c mice that had been housed individually except when otherwise indicated. Mice were subjected to the various treatments for 1.5 h between 7:00 and 11:00. Bedding that had been soiled for two days was used for stimulation. For time course experiments, soiled bedding from six adult male mice was collected, mixed and divided into six equal parts that were then used to stimulate mice for six different periods of time. For male urine-soiled bedding, 1.5 ml male urine was applied to 10 g of clean bedding. For testing of putative pheromones, 30 g bedding was wetted with 1.5 ml of 0.1% DMSO containing 1 mM 2-heptanone, 0.5 mM DHB or 0.5 mM SBT. For exposure to fur, 35–40 mg fur was collected from the head and anogenital regions of male mice and placed on bedding. For stimulation with an extract of exocrine glands, each gland was homogenized in 10 ml of 20 mM Tris-HCl (pH 7.5) per g of tissue. The extract was centrifuged, and 35–40 mg of a cotton swab was soaked in supernatant equivalent to one gland. The cotton swab was then dried in a Speed Vac and placed on bedding.

For analysis of ELG column fractions, cotton swabs were soaked with aliquots equivalent to five ELG glands from individual DEAE column fractions, and then

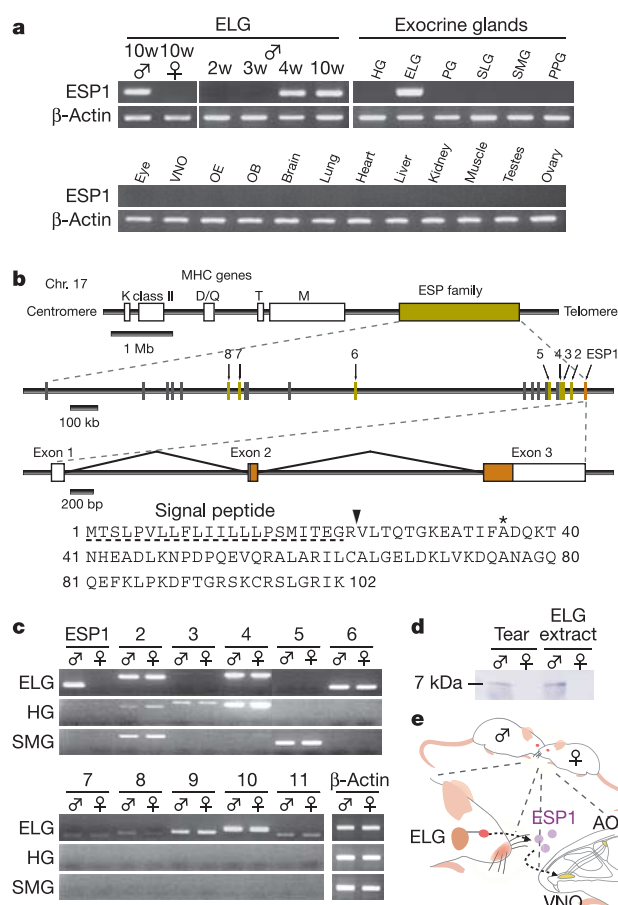


Figure 4 | Analysis of genes encoding ESP1 and its homologues.

a, Expression pattern of the ESP1 gene at different ages (2–10 weeks) and in different tissues, as determined by RT-PCR. OE, olfactory epithelium; OB, olfactory bulb. **b**, Genomic structure and map of ESP1 and 20 other homologues. Numbers were assigned to members of the ESP gene family expressed in any of the six exocrine glands. The locations of three homologues (ESP9–11) have not been determined in NCBI Mouse Build 33. The deduced amino-acid sequence of the ESP1 gene is shown. Dark orange, the coding region of the ESP1 gene; arrowhead, the boundary between exon 2 and 3; asterisk, the beginning of the N terminus of purified ESP1. **c**, Expression profile of the ESP gene family as determined by RT-PCR. **d**, Secretion of ESP1 in male tear fluid, measured by western blot. **e**, A model of ESP1-mediated sexual communication. Male mice produce ESP1 in the extraorbital lacrimal gland, which is secreted from the eyes to the outside. The secreted ESP1 is transferred to the female VNO, resulting in recognition of the ESP1 signal through the vomeronasal pathway.

assayed as described for the analysis of exocrine gland extracts. For analysis of C4 column fractions, a sample equivalent to five ELG glands was lyophilized and dissolved in 200 μ l of 20 mM Tris-HCl (pH 7.5) before applying it to a cotton swab, drying it, and assaying it as described for the analysis of exocrine gland extracts. For this analysis, it was necessary to re-dissolve the sample in Tris-HCl buffer to ensure activity.

Immunohistochemistry. Double-label immunohistochemistry with antibodies against c-Fos and $G\alpha_o$ was carried out as described previously²³. Additional information is provided in Supplementary Methods.

In situ hybridization. The Ensembl mouse genome database v24.33.1 (<http://www.ensembl.org/>) predicted 88 V2R genes, 44 of which were annotated to encode full-length V2Rs. Twenty sets of specific primers were designed from the lowest homologous region of V2Rs, and were expected to amplify 44 V2R genes. Polymerase chain reaction with reverse transcription (RT-PCR) resulted in amplification of V2R genes from the VNO using 12 out of the 20 primer sets. These 12 different V2R complementary DNAs (each of which is ~600 bp, including exon 4 and a part of exons 3 and 5) were used to generate antisense digoxigenin-labelled RNA probes. On the basis of homology (>80% or <80%), these probes were expected to hybridize with a total of 36 V2Rs. The VR4 probe was predicted to hybridize with seven V2R genes, including V2r4, V2r5, V2r14 (ref. 13) and EC2-V2R (ref. 8), and the V2R1 probe to three V2Rs, including V2r1b (ref. 14). The V2Ro probe was used for detecting a single V2R. The V2Rp probe hybridizes with two annotated V2R genes (V2Rp1, V2Rp2) and several other non-annotated V2R genes that are highly homologous in the database. Additional information is provided in the Supplementary Methods.

Purification and characterization of active peptides. ELG glands from BALB/c adult male mice were homogenized in 10 ml of 20 mM Tris-HCl (pH 7.5) per g tissue. After centrifugation, the supernatant was dialysed against 20 mM Tris-HCl (pH 7.5). The dialysed fluid was passed through a 0.45- μ m MILLEX-HV filter (Millipore) and loaded onto an anion-exchange high-performance liquid chromatography (HPLC) column (TSK-GEL DEAE-5PW, Tosoh). A gradient of 0–0.5 M NaCl in 20 mM Tris-HCl (pH 7.5) was applied at 0.5 ml min⁻¹, and the flow-through and 2-ml fractions were collected and then assayed. The flow-through sample containing the activity was lyophilized, dissolved in 10% acetonitrile (ACN) in 0.1% trifluoroacetic acid (TFA), and loaded onto a reverse-phase HPLC column (PEGASIL-300 C4P, Senshu). Bound peptides were eluted with a gradient of 10–60% ACN in 0.1% TFA at 1 ml min⁻¹, and 1-ml fractions were collected. The active fractions were lyophilized, the residues were dissolved in 30% ACN in 0.1% heptafluorobutyric acid (HFBA), and further separated using the same C4P column but using a gradient of 30–60% ACN in 0.1% HFBA at 1 ml min⁻¹. Individual ultraviolet 220-nm absorption peaks were collected, and each fraction was assayed. The two active peaks at fractions A and B were subjected to N-terminal protein sequencing (Prosize, Applied Biosystems). Molecular masses of peptides in the two fractions were determined by mass spectrometry (Voyager-DE STR, Applied Biosystems) and calibrated using bovine insulin as an external control. The location of the signal peptide cleavage site was predicted using the SignalP server (<http://www.cbs.dtu.dk/services/SignalP/>).

Electrophysiology. The VNO was dissected out from an adult female mouse and placed in a 2% agar bed containing solution A (140 mM NaCl, 5.6 mM KCl, 5 mM HEPES, 2 mM pyruvic acid sodium salt, 1.25 mM KH₂PO₄, 2 mM CaCl₂, 2 mM MgCl₂, 9.4 mM D-glucose, pH 7.4). The tissue was superfused with oxygen-impregnated solution A throughout the experiment. Slow transepithelial potential changes evoked by stimuli were recorded from the vomeronasal neuroepithelium using a glass pipette (0.5–2 M Ω tip resistance, filled with solution A and 2% agar) that was connected by an Ag/AgCl bridge to a DC-differential amplifier (ER-1, Cygnus). A second Ag/AgCl bridge was connected to another glass pipette inserted into the agar bed and served as a reference electrode. Electrical signals were band-passed at DC–100 Hz, stored on a Macintosh computer using a 1-kHz sampling recorder, and processed with a 20 Hz low-pass digital filter. A chemically synthesized MHC class I peptide (SYFPEITHI)²⁹ was purchased from Thermo Electron GmbH. All stimulants were diluted in solution A just before use. Each stimulus was applied focally to the luminal surface of the sensory epithelia through a barrelled glass capillary (70–120 μ m inner diameter) attached to a pressure ejection system.

Preparation of rESPI, RT-PCR analysis and western blot analysis. These procedures are described in the Supplementary Methods.

Received 10 February; accepted 11 July 2005.

- Dulac, C. & Torello, A. T. Molecular detection of pheromone signals in mammals: from genes to behaviour. *Nature Rev. Neurosci.* **4**, 551–562 (2003).
- Brennan, P. A. & Keverne, E. B. Something in the air? New insights into mammalian pheromones. *Curr. Biol.* **14**, R81–R89 (2004).

- Clancy, A. N., Coquelin, A., Macrides, F., Gorski, R. A. & Noble, E. P. Sexual behaviour and aggression in male mice: involvement of the vomeronasal system. *J. Neurosci.* **4**, 2222–2229 (1984).
- Wysocki, C. J. & Lepri, J. J. Consequences of removing the vomeronasal organ. *J. Steroid Biochem. Mol. Biol.* **39**, 661–669 (1991).
- Meredith, M. Chronic recording of vomeronasal pump activation in awake behaving hamsters. *Physiol. Behav.* **56**, 345–354 (1994).
- Stowers, L., Holy, T. E., Meister, M., Dulac, C. & Koentges, G. Loss of sex discrimination and male-male aggression in mice deficient for TRP2. *Science* **295**, 1493–1500 (2002).
- Leypold, B. G. et al. Altered sexual and social behaviors in trp2 mutant mice. *Proc. Natl Acad. Sci. USA* **99**, 6376–6381 (2002).
- Loconto, J. et al. Functional expression of murine V2R pheromone receptors involves selective association with the M10 and M1 families of MHC class Ib molecules. *Cell* **112**, 607–618 (2003).
- Norlin, E. M., Gussing, F. & Berghard, A. Vomeronasal phenotype and behavioural alterations in *Gai2* mutant mice. *Curr. Biol.* **13**, 1214–1219 (2003).
- Dulac, C. & Axel, R. A novel family of genes encoding putative pheromone receptors in mammals. *Cell* **83**, 195–206 (1995).
- Berghard, A. & Buck, L. B. Sensory transduction in vomeronasal neurons: evidence for $G\alpha_o$, *Gai2*, and adenylyl cyclase II as major components of a pheromone signalling cascade. *J. Neurosci.* **16**, 909–918 (1996).
- Herrada, G. & Dulac, C. A novel family of putative pheromone receptors in mammals with a topographically organized and sexually dimorphic distribution. *Cell* **90**, 763–773 (1997).
- Matsunami, H. & Buck, L. B. A multigene family encoding a diverse array of putative pheromone receptors in mammals. *Cell* **90**, 775–784 (1997).
- Ryba, N. J. & Tirindelli, R. A new multigene family of putative pheromone receptors. *Neuron* **19**, 371–379 (1997).
- Krieger, J. et al. Selective activation of G protein subtypes in the vomeronasal organ upon stimulation with urine-derived compounds. *J. Biol. Chem.* **274**, 4655–4662 (1999).
- Leinders-Zufall, T. et al. Ultrasensitive pheromone detection by mammalian vomeronasal neurons. *Nature* **405**, 792–796 (2000).
- Hurst, J. L. et al. Individual recognition in mice mediated by major urinary proteins. *Nature* **414**, 631–634 (2001).
- Boschat, C. et al. Pheromone detection mediated by a V1r vomeronasal receptor. *Nature Neurosci.* **5**, 1261–1262 (2002).
- Luo, M., Fee, M. S. & Katz, L. C. Encoding pheromonal signals in the accessory olfactory bulb of behaving mice. *Science* **299**, 1196–1201 (2003).
- Halem, H. A., Cherry, J. A. & Baum, M. J. Vomeronasal neuroepithelium and forebrain Fos responses to male pheromones in male and female mice. *J. Neurobiol.* **39**, 249–263 (1999).
- Kumar, A., Dudley, C. A. & Moss, R. L. Functional dichotomy within the vomeronasal system: distinct zones of neuronal activity in the accessory olfactory bulb correlate with sex-specific behaviors. *J. Neurosci.* **19**, RC32 (1999).
- Halem, H. A., Baum, M. J. & Cherry, J. A. Sex difference and steroid modulation of pheromone-induced immediate early genes in the two zones of the mouse accessory olfactory system. *J. Neurosci.* **21**, 2474–2480 (2001).
- Kimoto, H. & Touhara, K. Induction of c-Fos expression in mouse vomeronasal neurons by sex-specific non-volatile pheromone(s). *Chem. Senses* **30** (suppl. 1), i146–i147 (2005).
- Ishii, T., Hirota, J. & Mombaerts, P. Combinatorial coexpression of neural and immune multigene families in mouse vomeronasal sensory neurons. *Curr. Biol.* **13**, 394–400 (2003).
- Kikuyama, S. et al. Sodefrin: a female-attracting peptide pheromone in newt cloacal glands. *Science* **267**, 1643–1645 (1995).
- Wabnitz, P. A., Bowie, J. H., Tyler, M. J., Wallace, J. C. & Smith, B. P. Aquatic sex pheromone from a male tree frog. *Nature* **401**, 444–445 (1999).
- Feldhoff, R. C., Rollmann, S. M. & Houck, L. D. in *Advances in Chemical Signals in Vertebrates* 117–25 (Kluwer-Plenum, New York, 1999).
- Rollmann, S. M., Houck, L. D. & Feldhoff, R. C. Proteinaceous pheromone affecting female receptivity in a terrestrial salamander. *Science* **285**, 1907–1909 (1999).
- Leinders-Zufall, T. et al. MHC class I peptides as chemosensory signals in the vomeronasal organ. *Science* **306**, 1033–1037 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank K. Yamamoto for helping protein sequence analysis, K. Mori for providing DHB and SBT, and S. Sakuda and H. Kataoka for discussions. This work was supported in part by grants from the Japan Society for the Promotion of Science (JSPS) and the Program for Promotion of Basic Research Activities for Innovative Biosciences (PROBRAIN) of Japan.

Author Information cDNA sequences have been deposited in GenBank under accession numbers AB194091 (ESPI), AB194093 (V2Rp1), AB194094 (V2Rp2) and AB194095 (V2Ro). Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to K.T. (touhara@k.u-tokyo.ac.jp).

LETTERS

STIM1 is a Ca^{2+} sensor that activates CRAC channels and migrates from the Ca^{2+} store to the plasma membrane

Shenyuan L. Zhang¹, Ying Yu¹, Jack Roos², J. Ashot Kozak¹, Thomas J. Deerinck³, Mark H. Ellisman³, Kenneth A. Stauderman² & Michael D. Cahalan¹

As the sole Ca^{2+} entry mechanism in a variety of non-excitabile cells, store-operated calcium (SOC) influx is important in Ca^{2+} signalling and many other cellular processes^{1–3}. A calcium-release-activated calcium (CRAC) channel in T lymphocytes is the best-characterized SOC influx channel^{4–6} and is essential to the immune response, sustained activity of CRAC channels being required for gene expression and proliferation^{7–10}. The molecular identity and the gating mechanism of SOC and CRAC channels have remained elusive. Previously we identified *Stim* and the mammalian homologue STIM1 as essential components of CRAC channel activation in *Drosophila* S2 cells and human T lymphocytes¹¹. Here we show that the expression of EF-hand mutants of *Stim* or STIM1 activates CRAC channels constitutively without changing Ca^{2+} store content. By immunofluorescence, EM localization and surface biotinylation we show that STIM1 migrates from endoplasmic-reticulum-like sites to the plasma membrane upon depletion of the Ca^{2+} store. We propose that STIM1 functions as the missing link between Ca^{2+} store depletion and SOC influx, serving as a Ca^{2+} sensor that translocates upon store depletion to the plasma membrane to activate CRAC channels.

We previously characterized a SOC current in *Drosophila* S2 cells with biophysical properties similar to CRAC channels in human T cells¹². More recently, *Stim* was identified in an RNA-mediated interference (RNAi)-based screen with the SERCA (sarcolemmal/endoplasmic reticulum Ca^{2+} ATPase) pump inhibitor thapsigargin (TG) to evoke SOC influx in *Drosophila* S2 cells¹¹. Uniquely among the 170 candidate genes that were screened, including all *trp*-related genes, RNAi-mediated suppression of *Stim* inhibited the Ca^{2+} influx evoked by TG. By single-cell imaging of cytosolic Ca^{2+} concentration ($[\text{Ca}^{2+}]_i$) and patch clamp analysis, we confirmed a functional requirement for *Stim*, and for the human homologue STIM1, to mediate CRAC channel activity in S2 cells and in Jurkat T cells, respectively. *Drosophila* *Stim* and mammalian STIM1 (collectively referred to here as Stim1) are modular type I transmembrane proteins with an EF-hand motif near the amino terminus located in the lumen of the endoplasmic reticulum (ER) or outside the cell (Supplementary Fig. 1)¹³. Because Stim1 does not resemble any known ion channel, the presence of the EF-hand motif and its localization indicated that Stim1 might function as a sensor of the ER Ca^{2+} store. According to this proposal, Ca^{2+} binding to the EF-hand domain of Stim1 within the lumen of the Ca^{2+} store would keep CRAC channels in the plasma membrane closed.

To test this possibility, full-length STIM1 and *Stim* cDNA were cloned from RBL cells and S2 cells and placed into appropriate

expression vectors. Mutants in the EF-hand region were prepared, two for STIM1 and four for *Stim*, on residues known to be critical for Ca^{2+} binding^{14,15}. Overexpression of wild-type (WT) or mutant Stim1 after transient transfection was confirmed by western blotting (Supplementary Fig. 2). In single-cell Ca^{2+} imaging experiments, overexpression of WT Stim1 produced no significant difference in resting $[\text{Ca}^{2+}]_i$, TG-independent Ca^{2+} influx or TG-evoked store release compared with control blank-transfected cells (Jurkat cells are shown in Fig. 1, S2 cells in Supplementary Fig. 3). A modest increase in TG-dependent (store-operated) Ca^{2+} influx was seen in Jurkat T cells (Fig. 1b) but not in S2 cells, consistent with the hypothesis that Stim1 by itself is not a functional CRAC channel. In contrast, Jurkat or S2 cells transfected with the EF-hand mutants bore two severe phenotypes: resting $[\text{Ca}^{2+}]_i$ and TG-independent Ca^{2+} influx were increased from about 50 nM to more than 200 nM on average. Histograms of resting $[\text{Ca}^{2+}]_i$ show that expression of EF-hand mutants increased resting $[\text{Ca}^{2+}]_i$ to more than 600 nM in many individual cells (Supplementary Fig. 4). The Ca^{2+} release transient, obtained by adding TG in zero- Ca^{2+} solution, was not changed (Supplementary Fig. 5), showing that the Ca^{2+} store content was not affected. To test whether the high values of resting $[\text{Ca}^{2+}]_i$ and enhanced TG-independent Ca^{2+} influx were caused by constitutively opened CRAC channels, 2-aminoethyl diphenyl borate, SKF96365 and Gd^{3+} were applied as pharmacological tools that block CRAC channels in Jurkat cells and in S2 cells^{2,3,12,16–18}. Each of these agents inhibited TG-evoked SOC influx in control cells and, at the same concentrations, all three inhibited both the high resting Ca^{2+} concentration and the enhanced TG-independent Ca^{2+} influx in Jurkat and S2 cells transfected with EF-hand mutants (Fig. 1g–i, and Supplementary Fig. 3c). These results demonstrate that CRAC channels in Jurkat or S2 cells are constitutively opened by overexpression of Stim1 EF-hand mutants.

In addition to affecting resting $[\text{Ca}^{2+}]_i$, expression of *Stim* EF-hand mutants arrested the growth of S2 cells, whereas overexpressing WT *Stim* had a small effect compared with that in S2 cells undergoing blank transfection (Supplementary Fig. 6). Most S2 cells overexpressing *Stim* EF-hand mutants were stained by annexin V, an early marker for the exposure of phosphatidylserine that occurs during apoptosis¹⁹. The growth arrest and apoptosis were probably triggered by abnormally high resting $[\text{Ca}^{2+}]_i$ (refs 19–21). A similar but milder growth defect in Jurkat cells overexpressing STIM1 EF-hand mutants was also observed.

How does STIM1 regulate the activity of CRAC channels? The subcellular localization of STIM1 was examined by

¹Department of Physiology and Biophysics and Center for Immunology, University of California, Irvine, California 92697, USA. ²TorreyPines Therapeutics, Inc., La Jolla, California 92037, USA. ³National Center for Microscopy and Imaging Research, Center for Research in Biological Structure and the Department of Neurosciences, University of California, San Diego, La Jolla, California 92093, USA.

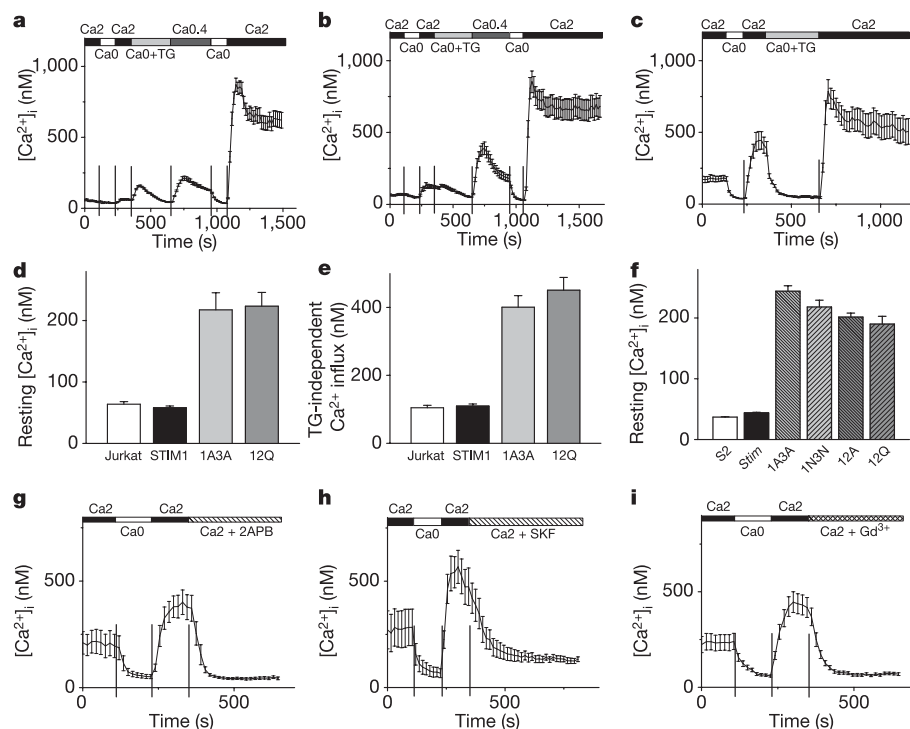


Figure 1 | Constitutive activation of CRAC channels by expression of EF-hand mutants of Stim or STIM1. **a–c**, Average $[Ca^{2+}]_i$ responses in blank-transfected Jurkat cells (**a**; $n = 59$), Jurkat cells overexpressing WT STIM1 (**b**; $n = 29$) and Jurkat cells overexpressing STIM1 EF12Q (**c**; $n = 18$). Cells were bathed in solutions as indicated: the number after 'Ca' reflects the Ca^{2+} concentration in mM. **d–f**, Resting $[Ca^{2+}]_i$ in Jurkat cells (**d**; from left to right: $n = 114$, $n = 171$, $n = 61$ and $n = 65$); TG-independent Ca^{2+} influx in Jurkat cells (**e**; $n = 114$, $n = 136$, $n = 61$ and $n = 42$); and resting $[Ca^{2+}]_i$ in S2 cells (**f**; $n = 392$, $n = 568$, $n = 425$, $n = 149$, $n = 316$ and $n = 102$). **g–i**, Effects of CRAC channel blockers 2-aminodiphenyl borate (2APB) (**g**, $50 \mu M$), SKF96365 (**h**, $20 \mu M$) and Gd^{3+} (**i**, $1 \mu M$) on TG-independent Ca^{2+} influx in Jurkat cells. The EF-hand mutants used were 12Q (**g**, $n = 16$), 12Q (**h**, $n = 8$) and 1A3A (**i**, $n = 28$). Error bars indicate s.e.m.

immunofluorescence staining of Jurkat cells overexpressing either WT STIM1 or STIM1 EF-hand mutants. In cells that overexpressed STIM1, most of the staining appeared within the ring of cytoplasm surrounding the nucleus, and very little was on the cell surface (Fig. 2a), which is consistent with a previous surface biotinylation study²². In contrast, EF-hand mutant STIM1 proteins that induce constitutive activation of CRAC channels were located predominantly on the surface membrane (Fig. 2b). Comparison with co-transfected green fluorescent protein (GFP) helped to define STIM1 or EF-hand mutant STIM1 at or below the cell surface. We reasoned that the altered localization of EF-hand mutants to the plasma membrane could result from their inability to bind Ca^{2+} within the intracellular store. Translocation of WT Stim1 from the Ca^{2+} store to the plasma membrane might therefore be a key event in the signalling pathway that activates CRAC channels. This hypothesis was supported by immunofluorescence staining of endogenous STIM1 in Jurkat T cells (Fig. 2c) and also in RBL cells, resting T cells from human donors and PC12 cells (Supplementary Figs 7 and 8). In cells pretreated with $1 \mu M$ TG in zero- Ca^{2+} external solution to deplete the Ca^{2+} store without causing a large increase in cytosolic $[Ca^{2+}]_i$, 'hotspots' of STIM1 staining were found at the cell surface, whereas a predominantly cytosolic distribution of STIM1 staining was seen in control cells bathed in Ringer solution containing $2 mM$ Ca^{2+} . In the control condition with Ca^{2+} stores full, STIM1 co-localized with ER markers (SERCA2 and protein disulphide isomerase) in Jurkat and in RBL cells, respectively, but after store depletion STIM1 accumulated in hotspots at the surface, and co-localization in the ER was less extensive (Fig. 2c, and Supplementary Fig. 8). STIM1 translocation kinetics were evaluated by fixing cells at different time points after treatment with TG or cyclopiazonic acid (CPA), another SERCA pump inhibitor that depletes ER Ca^{2+} stores and triggers SOC influx²³ (Fig. 2d and Supplementary Fig. 9a). When Ca^{2+} -store depletion was initiated, STIM1 moved rapidly to the surface membrane, reaching near-maximal surface expression within about 5 min and remaining on the surface for at least 40 min. For comparison, the inhibition of SERCA pumps activated CRAC currents with a somewhat more rapid time course in perforated patch recordings. Current begins to develop when STIM1 translocation is first observed, and

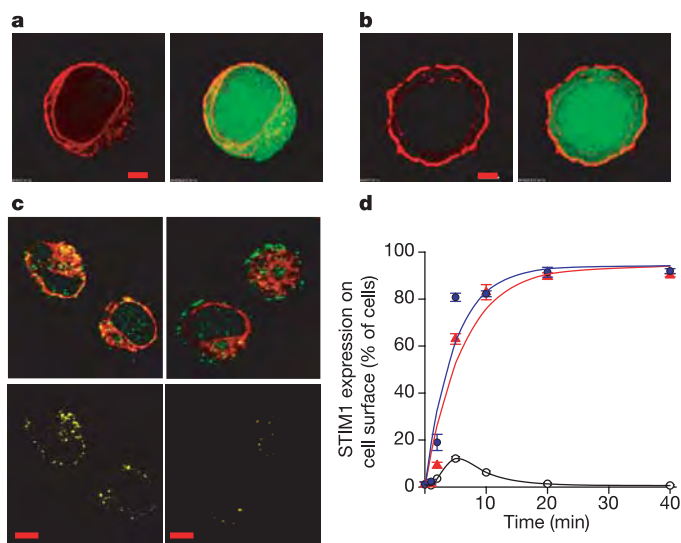


Figure 2 | Mutations in EF-hand motif or store depletion induce STIM1 translocation to the plasma membrane. **a, b**, STIM1 (red) immunofluorescence staining of Jurkat cells transfected with WT STIM1 (**a**) or EF1A3A STIM1 (**b**), in Ringer solution containing $2 mM$ Ca^{2+} (left). EGFP (green) was co-transfected to define the cytoplasmic region (right). Scale bar, $2 \mu m$. **c**, STIM1 (green) and SERCA2 (red) immunofluorescence staining of Jurkat cells in $2 mM$ Ca^{2+} Ringer solution (left) or $1 \mu M$ TG in zero- Ca^{2+} solution for 10 min (right). Note the separation of STIM1 at the cell surface (top right). Bottom: co-localization images (yellow) depicting pixels that contained both STIM1 and ER fluorescence. Note reduced co-localization caused by STIM1 translocation in TG-treated cells. Scale bar, $5 \mu m$. **d**, Time course of STIM1 translocation triggered by $1 \mu M$ TG (filled circles) or $10 \mu M$ CPA (triangles), represented by the percentage of cells with strong surface fluorescence. Open circles, $2 mM$ Ca^{2+} . An average of 315 cells were counted for each time point. Time constants: 4.7 min (TG), 6.1 min (CPA). Error bars indicate s.e.m.

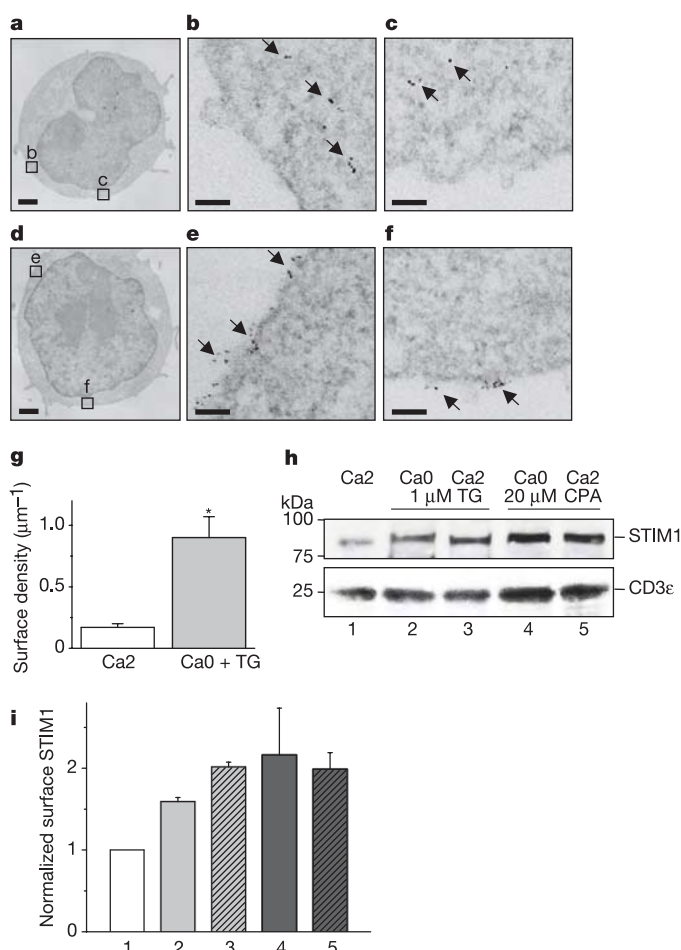


Figure 3 | Subcellular distribution of STIM1 before and after store depletion: immunoelectron microscopy and surface biotinylation.

a, Quantum-dot-labelled STIM1 in control Jurkat cells bathed in Ringer solution containing 2 mM Ca^{2+} . **b**, **c**, Enlargements of the boxed areas in **a**. **d**, STIM1 in Jurkat cells pretreated for 15 min with 1 μM TG in zero- Ca^{2+} solution. **e**, **f**, Enlarged regions of **d** showing clustered STIM1 distribution on the plasma membrane. Scale bar, 1 μm (**a**, **d**); 0.1 μm (**b**, **c**, **e**, **f**). **g**, STIM1 surface density, pooled data from control cells ($n = 14$) and store-depleted cells ($n = 13$); asterisk denotes $P < 0.01$ compared with controls. **h**, Jurkat cells were treated for 30 min as indicated. Biotinylated proteins were detected by antibodies against STIM1 (top) and CD3ε (bottom). **i**, Quantification of biotinylation experiments: there was a 1.6–2.2-fold increase upon store depletion. Error bars indicate s.e.m.

reaches a peak at 160 ± 66 s (mean $\pm$ s.d., $n = 6$ cells; Supplementary Fig. 8b), when STIM1 translocation rate is maximal. Additional experiments, summarized in Supplementary Fig. 9b–d, indicated that STIM1 surface translocation also occurred when $[\text{Ca}^{2+}]_i$ was elevated accompanying store depletion by treatment with TG in 2 mM Ca^{2+} external solution, but did not occur in cells treated with zero- Ca^{2+} solution in the absence of TG or in cells treated with 2 mM Ca^{2+} plus dimethylsulphoxide used to dissolve TG. Thus, depletion of intracellular Ca^{2+} stores is sufficient to trigger translocation of STIM1 proteins to the plasma membrane, indicating a possible mechanistic link to CRAC channel activation.

The translocation of STIM1 was further confirmed by immunoelectron microscopy, using quantum-dot-conjugated second antibodies after polyclonal anti-STIM1 antibodies to reveal STIM1 protein with single-molecule resolution. In control cells bathed in Ringer solution containing 2 mM Ca^{2+} , most quantum dots were located in the lumen of smooth ER-like vesicular structures within the cytosol and occasionally at the cell surface (Fig. 3a–c). In contrast,

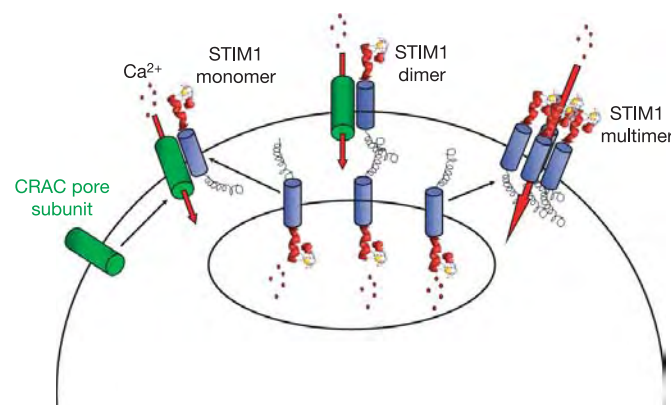


Figure 4 | Models of STIM1 function. Upon store depletion, STIM1 located in the ER unbinds Ca^{2+} and translocates to the plasma membrane to activate CRAC channel subunits that are already in the plasma membrane (left), form junctions between the Ca^{2+} store and the plasma membrane (middle) or assemble to form functional CRAC channels (right).

in TG-treated cells quantum dots appeared in clusters on the plasma membrane (Fig. 3d–f). The STIM1 surface density, with or without Ca^{2+} -store depletion, was determined by counting the number of quantum dots normalized to the linear distance along the plasma membrane. A fivefold increase in STIM1 surface density was obtained after treatment with TG to induce store depletion (Fig. 3g). Moreover, an increase in surface-accessible STIM1 was also detected by biotinylation (Fig. 3h, i).

The results can be summarized in two main conclusions: first, CRAC channels are activated when Stim1 is unable to bind Ca^{2+} within the store, either after Ca^{2+} release or if the EF-hand mutant is expressed; and second, Stim1 translocates to the plasma membrane upon Ca^{2+} store depletion. Previously we showed that STIM1 also regulates SOC influx in HEK-293 and SH-SY5Y cells¹¹. In addition, while this paper was in review, a report from Liou *et al.* showed that STIM1 has a function in the regulation of SOC influx in HeLa cells²⁴. This group also reported redistribution (but not membrane insertion) of YFP-tagged STIM1 upon store depletion, but did not examine endogenous STIM1. By showing translocation and membrane insertion of endogenous STIM1, our data suggest a crucial link to the functional role of STIM1. On the basis of these findings we propose a calcium sensor model to account for the role of Stim1 in regulating SOC influx. The proposed model is a modification of a vesicle fusion hypothesis for CRAC channel activation, in which Stim1 is the element that translocates. As illustrated in Fig. 4, Stim1 is initially located in the membrane of the Ca^{2+} store, with its low-affinity EF hand inside the lumen sensing Ca^{2+} and stabilized by bound Ca^{2+} when the store is full. When the store is depleted by inositol-1,4,5-trisphosphate-induced Ca^{2+} release or by inhibition of the SERCA pump, Ca^{2+} dissociates and Stim1 moves to the plasma membrane, presumably in vesicles that may insert into the membrane. Mutation of the EF hand mimics Ca^{2+} store depletion, initiating the translocation and activation of CRAC channels. Once Stim1 is at the cell surface it can activate the CRAC channel in one of three ways: it can interact with a putative pore-forming subunit, it can activate Ca^{2+} influx by means of conformational coupling through its coiled-coil domain, or it can assemble with additional Stim1 monomers and perhaps other components to form a unique functional Ca^{2+} channel. In addition to mechanistic insights, Stim1 may provide a biochemical handle with which to identify additional components of the translocation mechanism and the CRAC channel pore.

METHODS

Cell culture and transfection. *Drosophila* S2 cells (Invitrogen) were propagated at 27 °C in Schneider's medium (Invitrogen) supplemented with 12.5% FCS and 1% glutamine. Cells were seeded at a density of 10^6 cells ml^{-1} and passed when

the cells achieved a density of about 6×10^6 cells ml^{-1} . Jurkat T cells (ATCC) were maintained and propagated as detailed by the ATCC. RBL-2H3 cells and PC12 cells were cultured in Eagle's MEM and DMEM, respectively, supplemented with 10% fetal bovine serum in 5% CO_2 -humidified atmosphere at 37 °C. Cells were passed twice weekly. Human peripheral T lymphocytes were isolated from the blood of healthy volunteers with the use of CD3⁺ RosetteSep (StemCell Technologies) and Histopaque 1077 (Sigma) and cultured in RPMI supplemented with 10% fetal bovine serum. S2 cells and Jurkat cells were transfected (see clones described in Supplementary Information) using a Nucleofector (Amaxa) in accordance with the manufacturer's protocol. Forty-eight hours after transfection, protein expression was confirmed by western blotting, and cells were used for $[\text{Ca}^{2+}]_i$ imaging, immunocytochemistry and apoptosis assays.

Antibodies and western blotting. Cell extracts were prepared by washing the cells with PBS and then extracting proteins with lysis buffer (in mM): 10 Tris, 3 CaCl_2 , 2 MgCl_2 , 2.5% Nonidet P40, pH 7.5. Protein concentration was determined with the Pierce BCA Protein Reagent Kit. The extract was prepared for SDS-PAGE analysis by the addition of 4 × sample buffer (Invitrogen). Samples were resolved by SDS-PAGE and analysed by standard western blotting techniques. STIM1 polyclonal antibodies against a carboxy-terminal peptide (STIM1-CT, DNGSIGETDSSPGRKKFPLKIFKKPLKK) were used at a dilution of 1:2500. Monoclonal antibodies against glyceraldehyde3phosphate dehydrogenase were from Research Diagnostics and used at a dilution of 1:5000. Anti-V5/horseradish peroxidase monoclonal antibodies were from Invitrogen and used at a dilution of 1:5000. Proteins were detected by developing with the SuperSignal (Pierce) detection system.

Single-cell $[\text{Ca}^{2+}]_i$ imaging. Ratiometric $[\text{Ca}^{2+}]_i$ imaging experiments were performed on S2 and Jurkat cells with Fura-2 as described¹¹, using solution recipes indicated in Supplementary Table 1. Transfected cells were recognized by co-expressed enhanced green fluorescent protein (EGFP), using filters to avoid contamination of Fura-2 fluorescence by bleed-through of GFP fluorescence²⁵. Data were analysed with Metafluor software (Universal Imaging) and OriginPro 7.5 software (OriginLab) and are expressed as means $\pm$ s.e.m.

Immunocytochemistry, light and electron microscopy. Stained cells were viewed under a confocal laser scanning microscope LSM510 META (Zeiss) or DeltaVision RT Restoration Imaging System. Q-dot-labelled cells were fixed for 10 min in 2% glutaraldehyde and postfixed for 20 min in 1% osmium tetroxide. Cells were then washed in distilled water, then dehydrated in ethanol and embedded in Durcupan ACM epoxy resin and subsequently prepared and viewed by electron microscopy (Jeol 2000FX).

Surface biotinylation. Cell-surface STIM1 protein was detected by cell-surface biotinylation and purification, in accordance with the manufacturer's instructions (Pierce), by using a modified protocol²² as described in Supplementary Information. Data were quantified with ImageJ software (NIH) and normalized as means $\pm$ s.e.m.

Perforated patch recording. CRAC current was activated by CPA or TG in RBL cells during perforated-patch recording with amphotericin B (0.2 mg ml^{-1}) in the pipette. See Supplementary Information for pipette and external solutions and recording protocols.

Received 22 April; accepted 22 August 2005.

- Putney, J. W. Jr, Broad, L. M., Braun, F. J., Lievremon, J. P. & Bird, G. S. Mechanisms of capacitative calcium entry. *J. Cell Sci.* **114**, 2223–2229 (2001).
- Putney, J. W. Jr Store-operated calcium channels: how do we measure them, and why do we care? [online]. *Sci. STKE* **2004** (243), pe37 (2004) (doi:10.1126/stke.2432004pe37).
- Lewis, R. S. Store-operated calcium channels. *Adv. Second Messenger Phosphoprotein Res.* **33**, 279–307 (1999).
- Parekh, A. B. & Putney, J. W. Jr Store-operated calcium channels. *Physiol. Rev.* **85**, 757–810 (2005).
- Lewis, R. S. & Cahalan, M. D. Mitogen-induced oscillations of cytosolic Ca^{2+} and transmembrane Ca^{2+} current in human leukemic T cells. *Cell Regul.* **1**, 99–112 (1989).
- Parekh, A. B. & Penner, R. Store depletion and calcium influx. *Physiol. Rev.* **77**, 901–930 (1997).

- Lewis, R. S. Calcium signalling mechanisms in T lymphocytes. *Annu. Rev. Immunol.* **19**, 497–521 (2001).
- Winslow, M. M., Neilson, J. R. & Crabtree, G. R. Calcium signalling in lymphocytes. *Curr. Opin. Immunol.* **15**, 299–307 (2003).
- Feske, S., Giltman, J., Dolmetsch, R., Staudt, L. M. & Rao, A. Gene regulation mediated by calcium signals in T lymphocytes. *Nature Immunol.* **2**, 316–324 (2001).
- Partiseti, M. *et al.* The calcium current activated by T cell receptor and store depletion in human lymphocytes is absent in a primary immunodeficiency. *J. Biol. Chem.* **269**, 32327–32335 (1994).
- Roos, J. *et al.* STIM1, an essential and conserved component of store-operated Ca^{2+} channel function. *J. Cell Biol.* **169**, 435–445 (2005).
- Yeromin, A. V., Roos, J., Stauderman, K. A. & Cahalan, M. D. A store-operated calcium channel in *Drosophila* S2 cells. *J. Gen. Physiol.* **123**, 167–182 (2004).
- Williams, R. T. *et al.* Identification and characterization of the STIM (stromal interaction molecule) gene family: coding for a novel class of transmembrane proteins. *Biochem. J.* **357**, 673–685 (2001).
- Beckingham, K. Use of site-directed mutations in the individual Ca^{2+} -binding sites of calmodulin to examine Ca^{2+} -induced conformational changes. *J. Biol. Chem.* **266**, 6027–6030 (1991).
- Nakayama, S. & Kretsinger, R. H. Evolution of the EF-hand family of proteins. *Annu. Rev. Biophys. Biomol. Struct.* **23**, 473–507 (1994).
- Ross, P. E. & Cahalan, M. D. Ca^{2+} influx pathways mediated by swelling or stores depletion in mouse thymocytes. *J. Gen. Physiol.* **106**, 415–444 (1995).
- Prakriya, M. & Lewis, R. S. Potentiation and inhibition of Ca^{2+} release-activated Ca^{2+} channels by 2-aminoethyl diphenyl borate (2-APB) occurs independently of IP_3 receptors. *J. Physiol. (Lond.)* **536**, 3–19 (2001).
- Chung, S. C., McDonald, T. V. & Gardner, P. Inhibition by SK&F 96365 of Ca^{2+} current, IL-2 production and activation in T lymphocytes. *Br. J. Pharmacol.* **113**, 861–868 (1994).
- Eray, M., Matto, M., Kaartinen, M., Andersson, L. & Pelkonen, J. Flow cytometric analysis of apoptotic subpopulations with a combination of annexin V-FITC, propidium iodide, and SYTO 17. *Cytometry* **43**, 134–142 (2001).
- Graber, M. N., Alfonso, A. & Gill, D. L. Ca^{2+} pools and cell growth: arachidonic acid induces recovery of cells growth-arrested by Ca^{2+} pool depletion. *J. Biol. Chem.* **271**, 883–888 (1996).
- Ferrari, D. *et al.* Endoplasmic reticulum, Bcl-2 and Ca^{2+} handling in apoptosis. *Cell Calcium* **32**, 413–420 (2002).
- Manji, S. S. *et al.* STIM1: a novel phosphoprotein located at the cell surface. *Biochim. Biophys. Acta* **1481**, 147–155 (2000).
- Dolmetsch, R. E. & Lewis, R. S. Signaling between intracellular Ca^{2+} stores and depletion-activated Ca^{2+} channels generates $[\text{Ca}^{2+}]_i$ oscillations in T lymphocytes. *J. Gen. Physiol.* **103**, 365–388 (1994).
- Liou, J. *et al.* STIM is a Ca^{2+} sensor essential for Ca^{2+} -store-depletion-triggered Ca^{2+} influx. *Curr. Biol.* **15**, 1235–1241 (2005).
- Fanger, C. M. *et al.* Calcium-activated potassium channels sustain calcium signalling in T lymphocytes. Selective blockers and manipulated channel expression levels. *J. Biol. Chem.* **276**, 12249–12256 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank L. Forrest for assistance with cell culture; A. Yeromin, O. Safrina and S. Wei for help with $[\text{Ca}^{2+}]_i$ imaging; G. Chandy for the use of molecular reagents and laboratory facilities; C. Hughes for providing access to the Amaxa Nucleofector; A. Kolski-Andreaco for the gift of pAc5.1/EGFP; K. Knowlton and D. Summers-Torres from the Department of Medicine, University of California, San Diego for help with deconvolution immunofluorescence microscopy; and P. J. DiGregorio, G. Velicelebi, M. Lioudyno, J. Hall and Y. Li for discussion. Confocal microscopy was performed at the Optical Biology Shared Resource, supported by a Developmental Biology Center and Cancer Center Support Grant at the University of California, Irvine. This work was supported by grants from the National Institutes of Health (to M.D.C. and M.H.E.) and by a fellowship from the Pulmonary Hypertension Association (to Y.Y.).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.D.C. (mcahalan@uci.edu).

LETTERS

Apolipoprotein-mediated pathways of lipid antigen presentation

Peter van den Elzen¹, Salil Garg², Luis León², Manfred Brigl², Elizabeth A. Leadbetter², Jenny E. Gumperz², Chris C. Dascher², Tan-Yun Cheng², Frank M. Sacks⁴, Petr A. Illarionov⁵, Gurdyal S. Besra⁵, Sally C. Kent³, D. Branch Moody² & Michael B. Brenner²

Peptide antigens are presented to T cells by major histocompatibility complex (MHC) molecules, with endogenous peptides presented by MHC class I and exogenous peptides presented by MHC class II. In contrast to the MHC system, CD1 molecules bind lipid antigens that are presented at the antigen-presenting cell (APC) surface to lipid antigen-reactive T cells¹. Because CD1 molecules survey endocytic compartments², it is self-evident that they encounter antigens from extracellular sources. However, the mechanisms of exogenous lipid antigen delivery to CD1-antigen-loading compartments are not known. Serum apolipoproteins are mediators of extracellular lipid transport for metabolic needs³. Here we define the pathways mediating markedly efficient exogenous lipid antigen delivery by apolipoproteins to achieve T-cell activation. Apolipoprotein E binds lipid antigens and delivers them by receptor-mediated uptake into endosomal compartments containing CD1 in APCs. Apolipoprotein E mediates the presentation of serum-borne lipid antigens and can be secreted by APCs as a mechanism to survey the local environment to capture antigens or to transfer microbial lipids from infected cells to bystander APCs. Thus, the immune system has co-opted a component of lipid metabolism to develop immunological responses to lipid antigens.

Serum, lymph and interstitial fluid are the physiological contexts in which extracellular lipid antigens are encountered by CD1-bearing antigen presenting cells *in vivo*. Lipid-transport particles termed lipoproteins are a major constituent of these extracellular compartments³, yet their role in CD1 antigen presentation has not been investigated. To assess their contribution to the immune recognition of lipids, we first used an *in vitro* system in which human dendritic cells (DCs) were cultured with CD1-presented lipid antigens in human serum to measure their uptake and stimulatory capacity for T cells. We used a model CD1d-presented glycolipid, galactosyl(α 1-2)galactosyl ceramide (GGC) because this antigen strictly requires uptake and delivery to lysosomes where it is converted to the active antigenic form, α -galactosyl ceramide (α -GC)⁴. GGC was allowed to distribute into human serum, after which we separated the various serum fractions to resolve the major lipoprotein components of serum, namely very-low-density lipoprotein (VLDL), low-density lipoprotein (LDL) and high-density lipoprotein (HDL). The fractions were then assayed for their ability to stimulate α -GC-reactive, CD1d-dependent natural killer (NK) T cells. The antigenic activity of GGC was distributed almost entirely in the VLDL fraction of human serum (Fig. 1a). VLDL is a complex of lipids and apolipoproteins, including apolipoprotein B (apoB) and apolipoprotein E (apoE), which independently participate in receptor-mediated uptake into cells^{5,6}. Only antibodies against apoE effectively blocked the

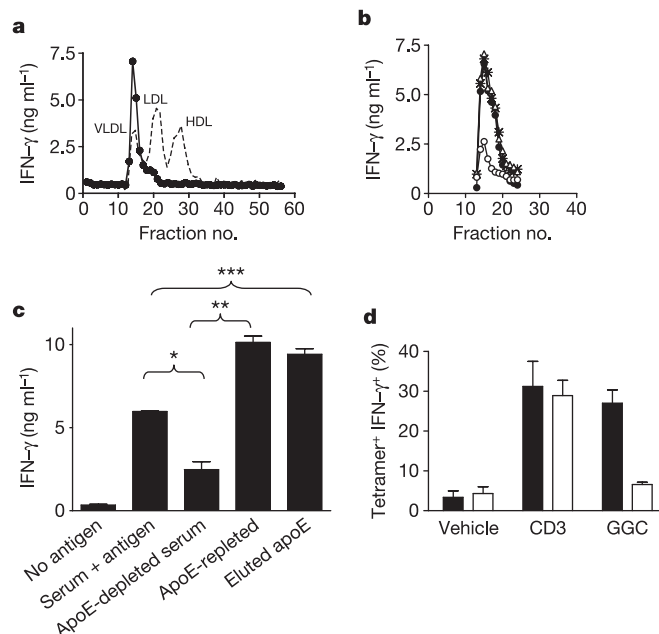


Figure 1 | Distribution and presentation of lipid antigens in serum, and dependence on apoE. **a**, VLDL localization of CD1-restricted T cell activity. GGC was incubated in human serum and separated by FPLC, and fractions were analysed for NK T-cell reactivity (solid line) and for cholesterol and triglycerides (dashed line). Data are representative of six separate experiments. IFN- γ , interferon- γ . **b**, Dependence of antigen reactivity on apoE. As in **a**, T-cell activity was measured in the absence (filled circles) or presence (open circles) of polyclonal blocking antibodies against apolipoprotein E versus isotype-matched antibodies against apolipoprotein B (triangles) and apolipoprotein A-II (asterisks). See also Supplementary Fig. 1. **c**, Depletion of apoE from serum diminishes GGC reactivity. T-cell reactivity to GGC was compared in 5% normal human serum, the same serum depleted of apoE, the apoE-depleted fraction plus added apoE (2.5 μ g ml⁻¹) and the apoE-positive fraction eluted from the column. Data are representative of four independent experiments. Asterisk, $P < 0.001$; two asterisks, $P < 0.001$; three asterisks, $P < 0.01$. **d**, Diminished CD1-restricted T-cell activity in apoE^{-/-} mice. Mice were injected intravenously with 200 ng of GGC, 1.5 μ g of anti-CD3, or vehicle alone. CD1d-tetramer-positive cells were analysed for intracellular interferon- γ production by flow cytometry. Data are representative of three separate experiments with 15 apoE^{-/-} mice (open bars) and 15 controls (filled bars). Error bars represent s.e.m.

¹Department of Pathology, ²Department of Medicine, Division of Rheumatology, Immunology and Allergy, and ³Center for Neurologic Disease, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁴Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts 02115, USA. ⁵School of Biosciences, The University of Birmingham, Edgbaston, Birmingham B15 2TT, UK.

VLDL-mediated T-cell activity (70–80% inhibition; see Supplementary Information), in comparison with isotype-matched antibodies against apoB or apoA-II (Fig. 1b). This blockade was specific for VLDL-bound GGC and did not affect the presentation of free GGC or phytohaemagglutinin (PHA) controls (Supplementary Fig. 1). Thus, the exogenously supplied lipid antigen GGC binds in the VLDL fraction of serum, and is acquired by DCs in an apoE-dependent manner, processed, and then presented to T cells.

To show the importance of apoE in serum in the uptake and presentation of this lipid antigen, we depleted apoE from human serum with an immunoaffinity column⁷. ApoE-depleted serum incubated with GGC had a significantly lower capacity to stimulate T cells than normal serum, and the presenting capacity was recovered by adding back purified apoE (Fig. 1c). Furthermore, the apoE-positive fraction of serum had a greater stimulatory capacity than serum alone. These results indicate that apoE might be the major factor in human serum mediating the presentation of GGC to CD1d-reactive T cells.

To assess the importance of apoE in the presentation of lipid antigens *in vivo*, we analysed the CD1d-restricted response to GGC in apoE-deficient mice. Although there were comparable numbers of α -GC-reactive NK T cells in apoE-deficient and wild-type mice, the response to intravenously administered GGC was drastically reduced in the apoE-deficient mice (Fig. 1d). This was not due to a non-specific deficit of T cell reactivity, because intravenous administration of anti-CD3 antibodies activated CD1d- α -GC tetramer-labelled T cells similarly in both apoE-deficient mice and controls. Thus, apoE has a crucial function in the presentation of an exogenous lipid antigen *in vivo*. However, it has been shown that much higher doses of α -GC can still activate NK T cells in apoE-deficient mice^{8,9}, indicating that less efficient alternative pathways of exogenous lipid presentation also exist.

To examine directly the binding of apoE to lipid antigens such as

GGC, we incubated antigens with apoE and then separated free from bound lipid with a heparin affinity column. Whereas free lipid readily passed through the column (Fig. 2a, upper panel), apoE bound with high affinity. After a 10-min incubation of apoE with GGC at room temperature, all detectable lipid antigen remained bound to apoE, was retained by the column and could be eluted with NaCl (Fig. 2a, lower panel). Even in the presence of a 500-fold excess of albumin, which itself can non-specifically bind lipids¹⁰, more than 95% of the GGC still bound to apoE. ApoE therefore binds efficiently to a model CD1-presented lipid antigen.

Next we determined whether delivery of lipid antigens bound to apoE enhances CD1-restricted T-cell responses in comparison with free lipid antigen. The antigenic potency of GGC incubated with VLDL-derived apoE or recombinant apoE was increased almost 50-fold (Supplementary Information) compared with freshly sonicated free GGC or GGC delivered with other lipoprotein or serum fractions such as HDL (not shown), LDL or albumin (Fig. 2b). This effect was specific for lipid-dependent T-cell responses because there was no effect of apoE on an MHC class II restricted response to tetanus toxoid, an exogenous protein antigen requiring endocytic processing (Fig. 2c) or on control T-cell stimulators such as CD3, PHA or phorbol 12-myristate-13-acetate (PMA) plus ionomycin (Supplementary Fig. 3). Given the known affinity of apoE for a wide variety of lipid classes³, we proposed that the binding and delivery of lipid antigens might apply to many CD1-presented lipid antigens. T-cell responses to naturally occurring microbial lipid antigens were also enhanced in the presence of apoE for the presentation of glucose monomycolate (GMM), mycolic acid (CD1b-presented) and mannosyl-phosphomycoketide (CD1c-presented)¹¹ (Fig. 2d and data not shown). Taken together, these results establish that apoE specifically enhances T-cell responses to a variety of CD1-presented lipid antigens for multiple CD1 isoforms.

The efficient uptake of soluble antigens by professional APCs such

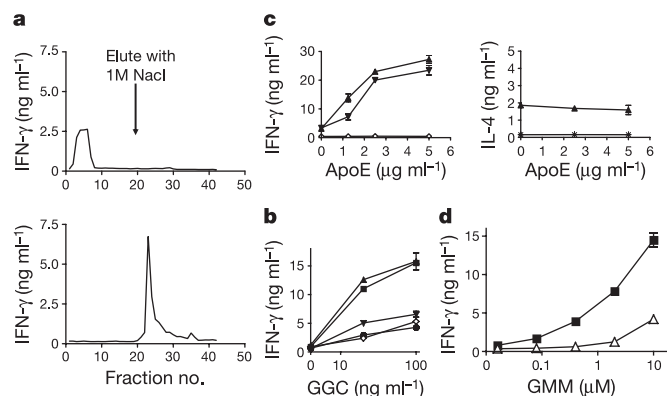


Figure 2 | ApoE binds lipid antigens and specifically enhances CD1-dependent T-cell responses. **a**, ApoE–GGC binding. Freshly sonicated GGC (top) or apoE-bound GGC (bottom) was loaded on a heparin affinity column and fractions were assayed for antigenic activity. ApoE was eluted with 1 M NaCl. Data are representative of at least six separate experiments. **b**, ApoE enhancement of GGC reactivity. GGC was incubated with apoE3 (squares), VLDL–apoE (upright triangles), LDL (inverted triangles) or BSA (circles), or freshly sonicated (diamonds), before being assayed for NK T-cell reactivity. **c**, ApoE specifically enhances CD1-dependent T-cell responses (left) but not MHC-dependent T-cell responses (right). α -GC (100 ng ml⁻¹; upright triangles in left panel), GGC (100 ng ml⁻¹; inverted triangles) or tetanus toxoid (TT, 10 μ g ml⁻¹; upright triangles in right panel) was incubated with various concentrations of apoE3 before being assayed for T-cell activity (left, α -GC-reactive clone; right, TT-reactive clone). IL, interleukin. **d**, Enhancement of a CD1b-dependent T-cell response against foreign antigens. GMM was incubated with or without apoE and assayed for antigenic activity with the GMM-reactive clone LDN5. All error bars represent s.e.m.

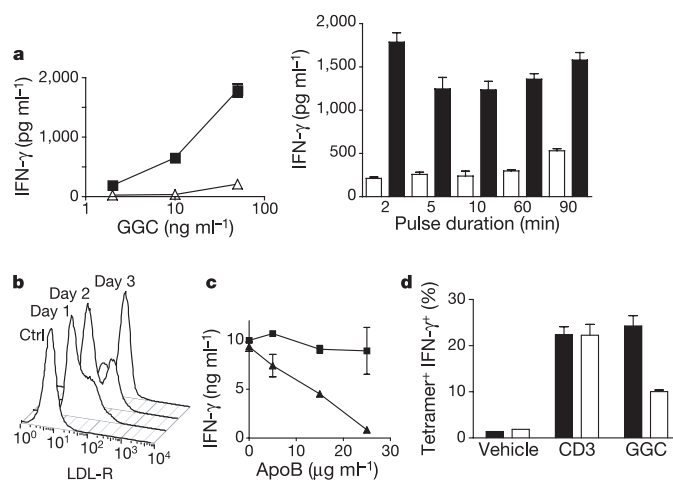


Figure 3 | Receptor-mediated uptake of apoE-bound lipid antigens. **a**, ApoE enhances the efficiency of uptake of GGC. ApoE-loaded GGC (filled squares (left) and filled bars (right)) or sonicated free GGC (open triangles (left) and open bars (right)) was used to pulse DCs for the indicated durations before being washed and tested for T-cell activity. Left: dose–response curve of GGC for a 2-min pulse. Right: various pulse durations are shown for 50 ng ml⁻¹ GGC. **b**, The LDL-R is upregulated as monocytes differentiate into DCs; monocyte-derived DCs were harvested on the indicated days and LDL-R expression was determined by flow cytometry compared with the control (Ctrl). **c**, Free GGC (squares) or apoE–GGC (triangles) was pulsed onto DCs in the presence of increasing concentrations of apoB, and NK T-cell activity was measured. Data are representative of at least three experiments. **d**, LDLR^{-/-} (open columns) and matched wild-type controls (filled columns) were challenged intravenously with 200 ng of GGC, 1.5 μ g of anti-CD3 or vehicle. Data are representative of two separate experiments with 10 LDLR^{-/-} and 10 wild-type mice.

as DCs is accomplished by robust macropinocytosis. In contrast, how APCs sample their local environment for lipid antigens is not known. ApoE is known to be secreted locally in tissues by macrophages and can account for 10% of all newly synthesized protein³. We also noted that apoE was coordinately upregulated with CD1 molecules in microarray experiments in human DCs^{12,13}, further indicating that apoE might be involved in lipid antigen presentation. We confirmed that apoE is abundantly produced by both DCs and macrophages, because apoE accumulated in supernatants at concentrations up to $10 \mu\text{g ml}^{-1}$ in culture (Supplementary Fig. 2). Thus, by locally producing apoE in their environment, DCs or macrophages can be able to promote the capture and presentation of lipid antigens to T cells.

ApoE is known to be taken up into the endocytic system of cells by receptor-mediated endocytosis¹⁴. Receptor-mediated uptake is much more efficient than macropinocytosis¹⁵ and may be of particular relevance for lipids. To determine the rapidity and efficiency of lipid uptake into DCs, we used free lipid antigen or apoE-bound lipid antigen to pulse DCs for varying durations before washing the cells and assaying their subsequent ability to present antigen to T cells. ApoE-delivered GGC achieved maximal stimulation of T cells after a

2-min exposure to DCs, whereas free lipid in medium alone required 4–6 h to achieve its maximal stimulation, which still reached only $48 \pm 11\%$ (mean $\pm$ s.e.m.) of that achieved in 2 min with apoE (Fig. 3a, and data not shown). This markedly enhanced rapidity and efficiency of uptake of the antigen is consistent with the difference between receptor-mediated endocytosis (apoE-delivered lipid) and macropinocytosis.

ApoE is known to be taken up by several cell-surface receptors, the most prominent being the LDL-R and the LDLR-like protein (LRP or CD91) (ref. 16). LRP is expressed on DCs¹⁷, but to our knowledge LDL-R has not been documented on DCs. Flow cytometric analysis showed surface expression of both LRP and LDL-R on DCs, the latter being markedly upregulated during the differentiation of monocytes to DCs (Fig. 3b). To determine the roles of these cell surface receptors in the uptake of apoE-bound lipid, we took complementary approaches. First we compared the ability of apoE3 (the most common human variant) with that of apoE2 (a variant with a single amino-acid change that diminishes binding to the LDL-R¹⁸) to mediate lipid antigen uptake in DCs. In comparison with apoE3, apoE2 had only 40–47% the capacity to enhance the CD1d-dependent response (data not shown), indicating that the LDL-R

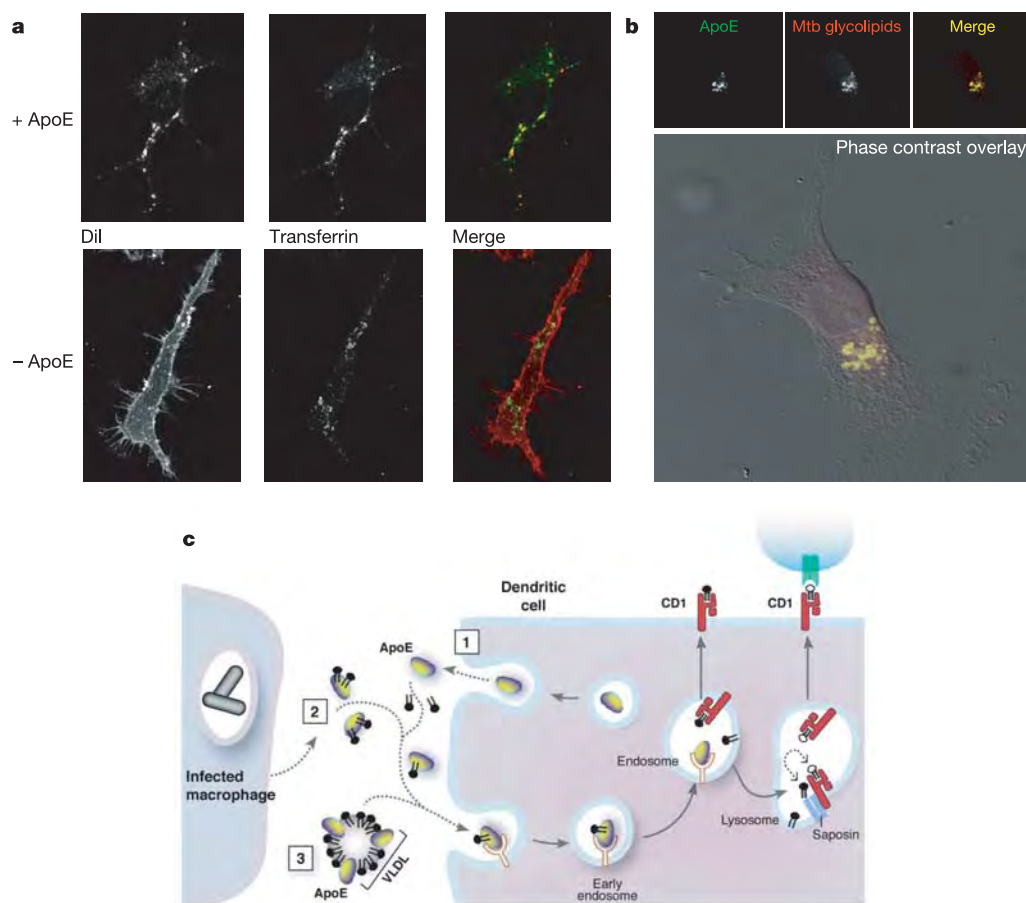


Figure 4 | Directed delivery of lipids to endosomal compartments by apoE. **a**, Dendritic cells were pulsed with DiI (red) with apoE (+apoE) or freshly sonicated (−apoE). Alexa-488 transferrin (green) was simultaneously pulsed with the lipid to reveal the early endosomes. After 5 min, cells were fixed and observed by confocal microscopy. Results are representative of multiple images in four separate experiments. **b**, ApoE transfers mycobacterial glycolipid antigens to uninfected DCs. *M. tuberculosis* (Mtb) was labelled with Alexa-555 and used to infect human macrophages. ApoE was isolated from culture supernatants by immunoaffinity purification and used to pulse DCs before observation by confocal microscopy. As controls, transfer of uninfected supernatants

showed apoE staining with no labelled glycolipid, and the infected apoE-negative fraction showed labelled glycolipid with no associated apoE (data not shown). **c**, Exogenous pathways for lipid antigen presentation mediated by apoE. Delivery of lipids to the endosomal system can occur by three possible mechanisms indicated by the numbers in squares: 1, secretion-capture, in which secreted (or recycled) apoE can capture lipid antigens; 2, bystander acquisition, in which infected cells such as macrophages may shed lipid antigens associated with apoE that are delivered to bystander (uninfected) DCs; and 3, serum lipoproteins (VLDL in the case of GGC) serve as a depot for lipid antigens and may stimulate a response far from the source of the antigen.

might have a significant function in the uptake of apoE–antigen complexes. Second, we took advantage of the fact that apoB, the primary component of LDL, also binds to the LDL-R and can serve as a competitive ligand but does not bind LRP¹⁹. DCs pulsed with apoE–GGC in the presence of increasing amounts of apoB exhibited a dose-dependent diminution of antigen presentation compared with controls (Fig. 3c and Supplementary Fig. 4) further implicating the LDL-R in uptake. Last, we analysed the response to GGC in LDL-R deficient mice. In comparison with wild-type controls, LDL-R knockouts had a greatly diminished response to antigen while responding similarly to non-specific stimulation by anti-CD3 (Fig. 3d). We therefore conclude that the LDL-R has a significant function in the uptake of apoE–lipid–antigen complexes and presentation by CD1.

Next, we wished to address the mechanism through which receptor-mediated uptake of apoE-bound lipids enhances presentation by CD1. We proposed that the apoE-mediated delivery of lipids resulted in more effective lipid antigen delivery to CD1 antigen loading compartments. To reveal the delivery of free lipids versus apoE-bound lipids directly, we used a fluorescent lipid probe, 1,1'-dioctadecyl-3,3,3',3'-tetramethylindocarbocyanine perchlorate (DiI). This molecule resembles CD1-presented lipid antigens in having two 18-carbon fatty-acyl chains and a distinct headgroup, in this case giving it fluorescent properties permitting detection within cells²⁰. The intracellular fate of DiI delivered as apoE-bound DiI was compared with that of free DiI by confocal microscopy in dendritic cells (Fig. 4a). ApoE-delivered DiI resulted in rapid uptake in a punctate endosomal pattern, co-localizing with the early endosomal marker transferrin (Fig. 4a, top panels). Freshly sonicated free DiI also rapidly entered dendritic cells but distributed in an apparent plasma membrane distribution with very little endosomal localization (Fig. 4a, bottom panels). Thus, receptor-mediated endocytosis of apoE concentrates lipid molecules rapidly in the endosomal system, allowing enhanced access to the trafficking pathways followed by CD1 (refs 1 and 2).

Last, we sought to reveal the apoE-mediated transfer of exogenous lipid antigens in the setting of an infection. Macrophages are the host cells during infection by *Mycobacterium tuberculosis*, harbouring the bacterium while being incapable of presenting lipid antigens by means of CD1a, CD1b or CD1c (owing to lack of expression). We therefore proposed that lipid antigens might be transferred to uninfected CD1a,b,c-positive DCs to enable antigen presentation. Because macrophages produce apoE abundantly, we infected these cells in culture with live *M. tuberculosis* labelled with a fluorescent dye to tag mycobacterial surface glycolipids²¹. The infected human macrophages were washed and cultured for two to three days, after which we isolated secreted apoE with an immunoaffinity column. The apoE fraction contained abundant fluorescently labelled bacterial glycolipid that was rapidly (in 15 min) taken up into DCs, accumulating in a punctate endosomal pattern completely co-localizing with apoE (Fig. 4b) and partly coinciding with CD1b (data not shown). ApoE produced by infected macrophages therefore transfers glycolipid antigens to uninfected 'bystander' DCs. It remains to be determined what mechanism transfers the lipid cargo from apoE to CD1 molecules.

These experiments outline an exogenous pathway for lipid antigen presentation mediated by apoE through receptor-mediated endocytosis. *In vivo*, the ability of bystander APCs to pick up microbial antigens from the extracellular fluid may be essential for effective immunity when infected cells cannot themselves present antigen. We propose that apoE readily binds exogenous lipid antigens and efficiently targets them for receptor-mediated uptake by DCs either in the infected tissue or in distant lymphoid organs to stimulate systemic immune responses. This may occur by the secretion and capture of apoE–lipid–antigen complexes in the local milieu or by apoE–VLDL–lipid–antigen complexes acquired from the circulation or from tissue fluid (Fig. 4c). We have not ruled out the role of other

pathways in CD1 antigen presentation, such as apoB/LDL-R²⁹ or oxidized LDL/scavenger receptors²², and consider it likely that structurally distinct lipid antigens may distribute differently in lipoprotein compartments. In addition to foreign lipids, apolipoproteins might also be important in providing a pathway for the delivery of self-lipid antigens and contribute to inflammatory diseases such as multiple sclerosis or atherosclerosis, in which both apoE and CD1 have been independently implicated^{9,16,23,24}. The exogenous lipid antigen delivery pathway mediated by apolipoproteins therefore has important implications for microbial immunity, autoimmunity and atherosclerosis.

METHODS

Cells and reagents. Immature monocyte-derived DCs were prepared from fresh leukapheresis mononuclear cells or peripheral blood mononuclear cells as described previously²⁵. Macrophages were generated by culturing leukocytes in IMDM medium for 7 days with 10% human serum and recombinant human macrophage colony-stimulating factor. The following CD1-restricted T cells and respective antigens were used as described previously: CD1d-restricted α -GC-reactive clones²⁵, CD1b-restricted T cell lines DN1 (mycolic acid-reactive)²⁶, and LDN5 (GMM-reactive)²⁷ and the CD1c-restricted mannose-phosphomycoketide-reactive T-cell line CD8-1 (ref. 11), tetanus-toxoid-specific MHC II-restricted clones²⁵.

Mice and analysis of NK T cell activation. ApoE-deficient mice and matched C57BL/6 controls were from Taconic. LDL-R-deficient mice and matched C57BL/6 controls were obtained from Jackson Laboratories. GGC was sonicated at $1 \mu\text{g ml}^{-1}$ in saline plus 0.045% Tween 20, and 200 μl was injected intravenously; alternatively, as controls, anti-CD3 ($1.5 \mu\text{g}$ per 200 μl) (BD Pharmingen) or vehicle alone was used. Spleens and livers were harvested after 2 h (for GGC) and 30 min (for anti-CD3) and analysed for the specific activation of NK T cells with the use of FITC-labelled α -GC-loaded CD1d tetramers, (TcR β -positive, CD19-negative) and intracellular staining for interferon- γ as described previously²⁵. Isotype-matched antibodies and unloaded CD1d tetramers were used as controls.

T-cell assays. Unless specified otherwise, DCs were pulsed with antigen (either apoE-loaded or freshly sonicated) in serum-free medium (SFM), consisting of RPMI supplemented with L-glutamine and penicillin/streptomycin and 4 mg ml^{-1} BSA (used as a non-specific carrier) or serum-free medium from Gibco. CD1-restricted T cells were then added directly (or following a wash step as in pulse-wash experiments; Fig. 3a) in the same SFM supplemented with human serum (final concentration 2%). Interferon- γ production by T cells was measured after an overnight (16–24 h) culture at 37°C . PHA (Sigma-Aldrich), anti-CD3 stimulation (antibody OKT3 at 100 ng ml^{-1}) and PMA/ionomycin were used as positive controls.

Lipid loading of apoE or other serum components. Lipid antigens were suspended in SFM by sonication in a water bath for 2 min. Recombinant human apoE3 and apoE2 (Invitrogen), VLDL-derived apoE and LDL-derived apolipoprotein B (Biodesign), HDL and LDL (Intracel) were then incubated at $5 \mu\text{g ml}^{-1}$ with lipid unless otherwise indicated.

FPLC separation of serum fractions by gel filtration. GGC was sonicated in DMSO at $100 \mu\text{g ml}^{-1}$, then added to 1 ml of human serum (filtered in a $5\text{-}\mu\text{m}$ filter (Pall Corporation)), incubated for 1–6 h at 37°C and loaded on a Superose-6 column (Amersham). Fractions (600 μl) were collected and assayed for T-cell reactivity as described above and for cholesterol and triglyceride with standard colorimetric assays (Thermo Electron Corporation) as described previously²⁸. Blocking antibodies were from Biodesign.

Depletion and isolation of apoE. Human AB serum (Gem Cell) was depleted of apoE as described previously⁷ with an immunoaffinity column. In brief, Econopac columns were packed with 2.5 ml of resin complex prepared from apoE polyclonal goat antibodies (Academy Biomedical) bound to Sepharose at 5.5 mg ml^{-1} . Depletion was confirmed by ELISA to be at least 95%. The apoE-positive fraction was eluted from the column with sodium thiocyanate, and both apoE-negative and positive fractions were washed with PBS and concentrated back to the original sample volume with 20-ml VivaSpin Concentrators (VivaScience).

Isolation of apoE-bound lipid by heparin column chromatography. Recombinant apoE ($5 \mu\text{g ml}^{-1}$) was incubated with lipid (100 ng ml^{-1} GGC or $5 \mu\text{g ml}^{-1}$ DiI) in loading buffer (10 mM NaH_2PO_4 pH 7). The sample was passed over 1 ml Hi-Trap Heparin columns (Amersham) and 250- μl fractions were collected in a 96-well plate (Costar). After being washed, the apoE was eluted with 1 M NaCl pH 7 and fractions were assayed for T-cell activity as described above.

Flow cytometry. Monocyte-derived DCs were prepared as described above and stained for flow cytometry with FITC-conjugated anti-LRP (CD91) antibodies (BD Pharmingen) and biotinylated rabbit anti-LDL-R (Research Diagnostics, Inc.). Streptavidin-Alexa-488 (Molecular Probes) was used as a secondary reagent (1:1000 dilution). Isotype-matched antibodies (BD Pharmingen) and secondary reagents were used as controls.

Confocal microscopy. DiI ($5 \mu\text{g ml}^{-1}$; Molecular Probes) was incubated for 1 h with apoE ($5 \mu\text{g ml}^{-1}$), and apoE-bound lipid was isolated as described above, with a heparin column. The isolated apoE-DiI complex was used to pulse live human dendritic cells adhered to fibronectin-coated coverslips. Free DiI ($5 \mu\text{g ml}^{-1}$) was freshly sonicated in SFM before pulsing DCs. Transferrin was conjugated to Alexa-488 (Molecular Probes) and used to pulse cells simultaneously at $5 \mu\text{g ml}^{-1}$. After a 5-min pulse, cells were fixed for 30 min with 2% paraformaldehyde and viewed by confocal microscopy on a Nikon TE2000 with EZ-C1 V.2.20 software. Images were generated with Adobe Photoshop.

Infection of macrophages with *M. tuberculosis*. *M. tuberculosis* (strain H37-Rv) was cultured in 7H9 medium with OADC supplements (Gibco) and 0.05% Tween 80. Mid-exponential phase cultures were labelled with Alexa-555-hydrazide (Molecular Probes) as described previously²¹ and used to infect human macrophages at a multiplicity of infection of 10. After a 1-h infection, macrophages were washed extensively to remove extracellular mycobacteria and cultured for three days in serum-free medium (Ex-Vivo medium; Gibco); supernatants were collected and passed through a $0.45\text{-}\mu\text{m}$ syringe filter to remove any whole bacteria or bacterial debris. apoE production was determined by ELISA, and apoE-lipid-antigen complexes were isolated with an immunoaffinity column as described above. The apoE-positive fraction (or, as controls, the apoE-negative or uninfected fractions) was used to pulse dendritic cells for 15 min and viewed by confocal microscopy as described above.

Statistical analysis. Unless otherwise stated, error bars and error values represent s.e.m.

Received 9 May; accepted 7 July 2005.

- Brigl, M. & Brenner, M. B. CD1: antigen presentation and T cell function. *Annu. Rev. Immunol.* **22**, 817–890 (2004).
- Sugita, M., Cernadas, M. & Brenner, M. B. New insights into pathways for CD1-mediated antigen presentation. *Curr. Opin. Immunol.* **16**, 90–95 (2004).
- Mahley, R. W., Weisgraber, K. H. & Farese, R. V. Jr in *Williams Textbook of Endocrinology* 10th edn (eds Larsen, P. R., Kronenberg, H. M., Melmed, S. & Polonsky, K. S.) 1642–1664 (Elsevier, Philadelphia, 2003).
- Prigozy, T. I. et al. Glycolipid antigen processing for presentation by CD1d molecules. *Science* **291**, 664–667 (2001).
- Krul, E. S., Tikkanen, M. J., Cole, T. G., Davie, J. M. & Schonfeld, G. Roles of apolipoproteins B and E in the cellular binding of very low density lipoproteins. *J. Clin. Invest.* **75**, 361–369 (1985).
- Sacks, F. M. & Krukonis, G. P. The influence of apolipoprotein E on the interactions between normal human very low density lipoproteins and U937 human macrophages: heterogeneity among persons. *Vasc. Med.* **1**, 9–18 (1996).
- Tomiyasu, K., Walsh, B. W., Ikewaki, K., Judge, H. & Sacks, F. M. Differential metabolism of human VLDL according to content of ApoE and ApoC-III. *Arterioscler. Thromb. Vasc. Biol.* **21**, 1494–1500 (2001).
- Major, A. S. et al. Quantitative and qualitative differences in proatherogenic NKT cells in apolipoprotein E-deficient mice. *Arterioscler. Thromb. Vasc. Biol.* **24**, 2351–2357 (2004).
- Tupin, E. et al. CD1d-dependent activation of NKT cells aggravates atherosclerosis. *J. Exp. Med.* **199**, 417–422 (2004).
- Rebbaa, A. & Portoukalian, J. Distribution of exogenously added gangliosides in serum proteins depends on the relative affinity of albumin and lipoproteins. *J. Lipid Res.* **36**, 564–572 (1995).
- Matsunaga, I. et al. *Mycobacterium tuberculosis* pks12 produces a novel polyketide presented by CD1c to T cells. *J. Exp. Med.* **200**, 1559–1569 (2004).
- Chaussabel, D. et al. Unique gene expression profiles of human macrophages and dendritic cells to phylogenetically distinct parasites. *Blood* **102**, 672–681 (2003).
- Le Naour, F. et al. Profiling changes in gene expression during differentiation and maturation of monocyte-derived dendritic cells using both oligonucleotide microarrays and proteomics. *J. Biol. Chem.* **276**, 17920–17931 (2001).
- Heeren, J. & Beisiegel, U. Intracellular metabolism of triglyceride-rich lipoproteins. *Curr. Opin. Lipidol.* **12**, 255–260 (2001).
- Lanzavecchia, A. Receptor-mediated antigen uptake and its effect on antigen presentation to class II-restricted T lymphocytes. *Annu. Rev. Immunol.* **8**, 773–793 (1990).
- Mahley, R. W. & Rall, S. C. Jr Apolipoprotein E: far more than a lipid transport protein. *Annu. Rev. Genomics Hum. Genet.* **1**, 507–537 (2000).
- Basu, S., Binder, R. J., Ramalingam, T. & Srivastava, P. K. CD91 is a common receptor for heat shock proteins gp96, hsp90, hsp70, and calreticulin. *Immunity* **14**, 303–313 (2001).
- Mahley, R. W. Apolipoprotein E: cholesterol transport protein with expanding role in cell biology. *Science* **240**, 622–630 (1988).
- Beisiegel, U., Weber, W., Ihrke, G., Herz, J. & Stanley, K. K. The LDL-receptor-related protein, LRP, is an apolipoprotein E-binding protein. *Nature* **341**, 162–164 (1989).
- Barak, L. S. & Webb, W. W. Fluorescent low density lipoprotein for observation of dynamics of individual receptor complexes on cultured human fibroblasts. *J. Cell Biol.* **90**, 595–604 (1981).
- Beatty, W. L. et al. Trafficking and release of mycobacterial lipids from infected macrophages. *Traffic* **1**, 235–247 (2000).
- Boullier, A. et al. Scavenger receptors, oxidized LDL, and atherosclerosis. *Ann. N.Y. Acad. Sci.* **947**, 214–222 (2001).
- Fazekas, F. et al. Apolipoprotein E epsilon 4 is associated with rapid progression of multiple sclerosis. *Neurology* **57**, 853–857 (2001).
- Shamshiev, A. et al. Self glycolipids as T-cell autoantigens. *Eur. J. Immunol.* **29**, 1667–1675 (1999).
- Brigl, M., Bry, L., Kent, S. C., Gumperz, J. E. & Brenner, M. B. Mechanism of CD1d-restricted natural killer T cell activation during microbial infection. *Nature Immunol.* **4**, 1230–1237 (2003).
- Porcelli, S., Morita, C. T. & Brenner, M. B. CD1b restricts the response of human CD4⁺ T lymphocytes to a microbial antigen. *Nature* **360**, 593–597 (1992).
- Moody, D. B. et al. Lipid length controls antigen entry into endosomal and nonendosomal pathways for CD1b presentation. *Nature Immunol.* **3**, 435–442 (2002).
- Innis-Whitehouse, W., Li, X., Brown, W. V. & Le, N. A. An efficient chromatographic system for lipoprotein fractionation using whole plasma. *J. Lipid Res.* **39**, 679–690 (1998).
- Brown, M. S. & Goldstein, J. L. A receptor mediated pathway for cholesterol homeostasis. *Science* **232**, 34–47 (1986).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank B. Asfaw, J. Heeren and S. Blacklow for helpful discussion, and D. Chausabel for providing additional microarray data. P.vdE., S.G., L.L., M. Brigl, E.L., J.G., C.C.D. and M. Brenner are funded by the NIH. L.L. is the recipient of a Howard Hughes Gilliam fellowship. Funding for T.-Y.C. and D.B.M. was from the Pew Foundation Scholars in the Biomedical Sciences, the Cancer Research Institute and the NIH. Funding for G.S.B., a Lister-Jenner Research Fellow, and for P.A.I. was from the Medical Research Council and the Wellcome Trust. S.C.K. was funded by the Boston Area Diabetes Research Center (BADERC).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.B. (mbrenner@rics.bwh.harvard.edu).

Chaperone release and unfolding of substrates in type III secretion

Yukihiro Akeda¹ & Jorge E. Galán¹

Type III protein secretion systems are essential virulence factors of many bacteria pathogenic to humans, animals and plants¹. These systems mediate the transfer of bacterial virulence proteins directly into the host cell cytoplasm. Proteins are thought to travel this pathway in a largely unfolded manner, and a family of customized cytoplasmic chaperones, which specifically bind cognate secreted proteins, are essential for secretion. Here we show that InvC, an ATPase associated with a *Salmonella enterica* type III secretion system², has a critical function in substrate recognition. Furthermore, InvC induces chaperone release from and unfolding of the cognate secreted protein in an ATP-dependent manner. Our results show a similarity between the mechanisms of substrate recognition by type III protein secretion systems and AAA+ ATPase disassembly machines.

Containing more than 20 proteins, type III secretion systems (TTSSs) are among the most complex protein secretion machines known^{1,3}. A central component of TTSSs is an envelope-associated organelle known as the needle complex⁴. This organelle is traversed by a central channel 28 Å in diameter, which is thought to serve as a conduit for the proteins transiting this secretion pathway⁵. Other essential components of TTSSs include an ATPase and a family of customized chaperones that specifically bind a ~100-amino-acid domain at the amino terminus of their cognate secretion substrates⁶. This domain wraps around homodimer chaperone pairs and is maintained in a non-globular conformation that retains significant secondary structure^{7,8}. After secretion, the chaperones must release their cognate substrate because they remain within the bacterial cytoplasm. In addition to the chaperone-binding domain, specific elements required for secretion have been also localized N-terminal of this domain of the secreted protein^{9–11}.

InvC is the ATPase associated with the TTSS encoded by *Salmonella enterica* serovar Typhimurium (*S. typhimurium*) within its pathogenicity island 1 (ref. 2), which mediates bacterial entry into mammalian cells and is essential for the virulence of this bacterial pathogen¹². Like other members of the TTSS-associated ATPase protein family¹³, InvC has extensive primary amino acid sequence similarity to the β -subunit of F_0F_1 -ATPases. Although these ATPases are essential for type III secretion (TTS), their specific role in this process is unknown. Their peripheral association to the cytoplasmic side of the bacterial membrane coupled to their ability to form hexameric rings^{14–16}, indicated the possibility that these ATPases might provide a link between the secretion substrates and the TTS machinery. We therefore investigated the potential role of InvC in TTS substrate recognition by examining its interaction with the type III secreted protein SptP bound to its cognate chaperone SicP. SptP is a SPI-1 TTS effector protein that poses two enzymatic activities: a Rho-family GTPase-activating protein (GAP) activity located immediately adjacent to the chaperone-binding domain, and a protein tyrosine phosphatase located at the carboxy terminus^{7,17,18}

(Fig. 1a). The full-length SptP (SptP^{1–543}) or its chaperone-binding domain (SptP^{1–158}) bound to its glutathione *S*-transferase (GST)-tagged cognate chaperone SicP were purified to homogeneity and their interaction with purified InvC was probed by affinity chromatography. The GST-SicP-SptP^{1–543} or GST-SicP-SptP^{1–158} complex specifically interacted with InvC but not with the irrelevant protein CdtA (Fig. 1b, d). The interaction occurred both in the presence of ATP and in the presence of ATP- γ S, indicating that the ability of InvC to hydrolyze ATP is not essential for substrate recognition (Fig. 1b). Both the monomeric and hexameric forms of InvC were able to interact with the SicP-SptP complex, although the latter appeared to do so more efficiently (Fig. 1c). InvC also interacted with the GST-tagged chaperone alone but did not interact with the effector domains of SptP (SptP^{161–543}) (Fig. 1d). These results indicate that InvC is able to recognize both the SicP chaperone and the chaperone/secreted protein complex, and that its multimeric form appears to be more efficient at substrate recognition. Another TTSS-associated ATPase has been also reported to bind TTSS chaperones¹⁹.

We have previously isolated loss-of-function mutations in InvC, which allowed the identification of specific residues important for its catalytic activity, membrane association, and oligomerization¹⁶. Mutations that affected either the membrane-binding (InvC^{V28M}) or catalytic activities (InvC^{K165E}) did not affect the ability of InvC to bind the SicP-SptP complex (Fig. 1e). However, a loss-of-function mutant (InvC^{L376P}) that retained its catalytic, membrane and oligomerization activities, did not bind the SicP-SptP complex (Fig. 1e). This mutant is located at the C terminus of InvC within a domain that has been predicted to face the cytoplasm¹⁶. Therefore, this mutant most likely defines the InvC substrate-binding domain. Taken together, these results indicate that InvC interacts with the chaperone and chaperone/effector protein complex through a domain located at its C terminus and therefore must play a critical role in substrate recognition by the TTS machine.

Immediately before secretion, TTSS-associated chaperones must dissociate from the chaperone/effector complex. We investigated whether the interaction of the SicP-SptP complex with InvC resulted in the release of the chaperone from its cognate substrate. Wild-type InvC or the catalytically inactive mutant InvC^{K165E} were incubated with purified GST-SicP-SptP^{1–158} complex in the presence of ATP or ATP- γ S and the dissociation of the complex was examined with an affinity chromatography assay. In the presence of wild-type InvC and ATP, SptP^{1–158} was released from its association with SicP (Fig. 2). Equivalent results were obtained with the GST-SicP-SptP^{1–543} complex (data not shown). In contrast, the release of SptP^{1–158} was not observed when the complex was incubated with the catalytically inactive InvC^{K165E} or in the presence of ATP- γ S (Fig. 2). These results indicate that InvC is capable of removing SptP from its chaperoned state and that this activity requires ATP hydrolysis.

¹Section of Microbial Pathogenesis, Yale University School of Medicine, New Haven, Connecticut 06536, USA.

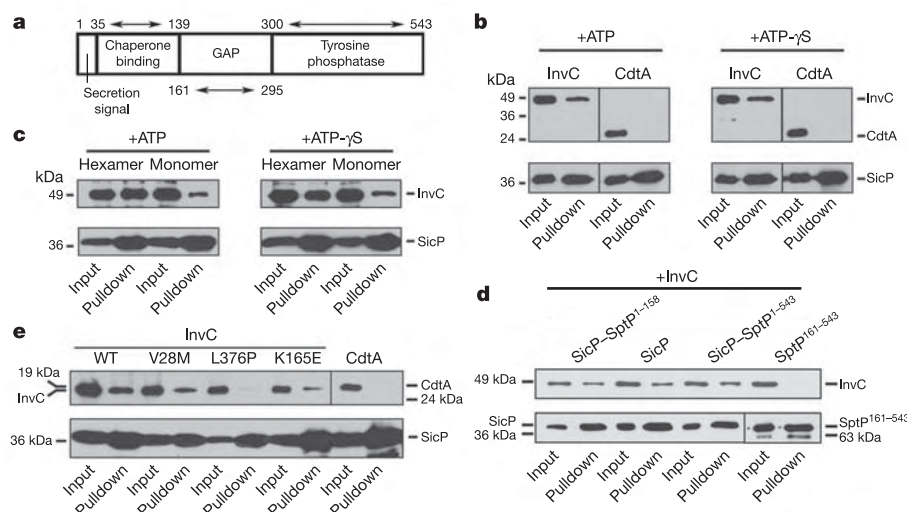


Figure 1 | The ATPase InvC interacts with TTS substrates. **a**, Diagram of the domain organization of the type III secreted protein SptP. **b**, Interaction of InvC with the SicP–SptP complex. Top, purified GST–SicP–SptP^{1–158} complex was incubated with InvC or CdtA (negative control) in the presence of ATP or ATP-γS. The presence of InvC or CdtA in the input sample (input) or bound to the GST–SicP–SptP^{1–158} complex (pulldown) was examined by western immunoblotting with an anti-His-tag monoclonal antibody. Bottom, the total amount of GST–SicP was determined by western

immunoblotting with an anti-GST antibody. **c**, The interaction GST–SicP–SptP^{1–158} with monomeric or hexameric InvC in the presence of ATP or ATP-γS was examined as described above. **d**, The interaction of InvC with GST–SicP, GST–SicP–SptP^{1–158}, GST–SicP–SptP^{1–543} or GST–SptP^{161–543} was examined as indicated above. **e**, The interaction of purified InvC mutants (as indicated) with the GST–SicP–SptP complex was examined as described above. WT, wild type.

The size constraints imposed by the diameter of the central channel of the TTSS-associated needle complex indicates that proteins travelling the TTS pathway must do so in a largely unfolded manner⁵. Although the TTSS-associated chaperones maintain the binding domains of their cognate secreted proteins in a non-globular state⁷, the enzymatic domains of the secreted proteins remain folded because they exhibit full activity²⁰. This observation indicated that unfolding of these domains must occur before secretion. We therefore examined whether InvC could mediate the unfolding of SptP upon its release from its cognate chaperone SicP. We first probed conformational changes in SptP by a protease susceptibility assay with the use of a monoclonal antibody that recognizes the C-terminal domain of SptP. Addition of InvC to the SicP–SptP complex in the presence of ATP resulted in a significant increase in the susceptibility of SptP to protease (Fig. 3a). In contrast, addition of InvC in the presence of ATP-γS or addition of the catalytically inactive mutant InvC^{K165E} did not increase the susceptibility of SptP to protease (Fig. 3a). These results are consistent with the hypothesis that InvC induces significant conformational changes in SptP and that this function requires its catalytic activity. We further explored this issue with a different assay. We reasoned that if InvC induces global unfolding of SptP, these changes should significantly affect its enzymatic activity. We therefore probed the conformational changes of SptP during its interaction with InvC by assaying its tyrosine phosphatase activity. The tyrosine phosphatase domain is the most distant from the chaperone-binding region, from which it is separated by the GAP domain (Fig. 1a). Addition of wild-type InvC to the SicP–SptP complex resulted in a drastic decrease in the phosphatase activity of SptP (Fig. 3b). In contrast, addition of the catalytically inactive InvC^{K165E} mutant had no effect on the phosphatase activity of SptP (Fig. 3b). Furthermore, InvC had no effect on the activity of SptP^{161–543}, a mutant that lacks the chaperone-binding domain (Fig. 3b), indicating that this domain is required for InvC to catalyse the unfolding of SptP. We also tested the unfolding of SptP by examining its 'transfer' to a GroEL 'trap' mutant (GroEL^{D87K}), which captures non-native proteins but is unable to release them²¹. We incubated the SicP–SptP^{1–543} complex with InvC or InvC^{K165E} in the presence of GroEL^{D87K} and subjected the mixture to gel-filtration chromatography. Addition of InvC, but not its catalytic mutant

InvC^{K165E}, significantly changed the migration of SptP (Supplementary Fig. 1). Instead of eluting at a position corresponding to the size of the SicP–SptP^{1–543} complex, SptP migrated with a broader profile and a significant proportion of the input migrated at the position of the 'trap', indicating the binding of non-native forms of SptP to GroEL^{D87K} (Supplementary Fig. 1). Addition of GroEL to the complex in the presence of InvC^{K165E} did not result in any change in the mobility of SptP, indicating that the GroEL trap alone does not affect the folding of SptP (Supplementary Fig. 1). Taken together, these results indicate that InvC can not only remove the chaperone but can also induce the global unfolding of SptP.

Secretion and translocation domains of type III secreted proteins are able to mediate the secretion of heterologous proteins²². However, secretion of some proteins is inefficient or does not occur at all²³, indicating that certain domains are not permissive for TTS because they cannot be unfolded by the ATPase. To test this hypothesis we fused the TTS targeting signals of SptP to the green fluorescent protein (GFP), which is known to have a very compact and stable fold. The GST–SicP–SptP^{1–158}–GFP protein complex was isolated and the ability of InvC to unfold the GFP domain was then examined

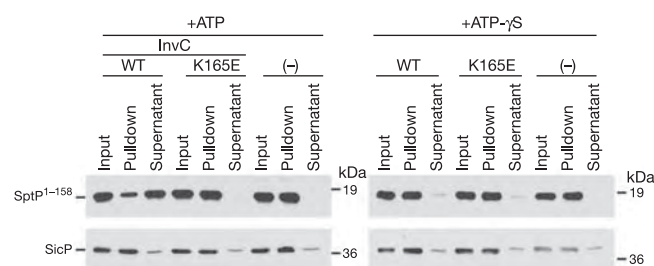


Figure 2 | InvC disassembles the chaperone/secreted protein complex. Purified InvC and the catalytic InvC^{K165E} mutant were added to purified GST–SicP–SptP^{1–158} complex in the presence of ATP or ATP-γS. The presence of released SptP^{1–158} (supernatant) or in association with GST–SicP (pulldown) was assayed by a GST-pulldown assay and western immunoblotting with a monoclonal antibody directed against the N terminus of SptP or the GST tag present in SicP.

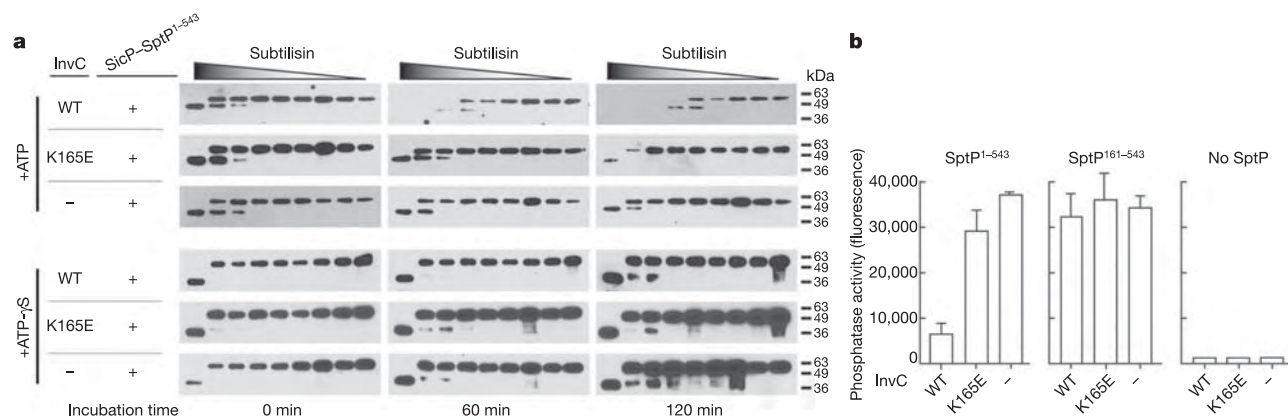


Figure 3 | InvC induces the unfolding of SptP. **a**, Purified SicP–SptP¹⁻⁵⁴³ complex was incubated with wild-type (WT) InvC or its catalytic mutant InvC^{K165E} in the presence of ATP or ATP-γS for the indicated durations. Decreasing amounts of subtilisin were then added and the presence of SptP in the different samples was analysed by western immunoblotting with an

antibody directed against its C terminus. **b**, WT InvC or its catalytic mutant InvC^{K165E} was added to the SicP–SptP complex or SptP¹⁶¹⁻⁵⁴³, and the tyrosine phosphatase activity was measured as indicated in Methods. Results are means ± s.d. for three independent measurements.

by monitoring the GFP fluorescence. Addition of InvC resulted in a release of SicP (Fig. 4a) but had no effect on the fluorescence activity of the SptP¹⁻¹⁵⁸–GFP fusion protein (Fig. 4b), indicating that InvC was unable to unfold the tightly packed GFP domain. If unfolding is a requirement for type III protein secretion, the inability of InvC to unfold GFP should prevent its secretion through the TTS pathway. We tested this hypothesis by examining the secretion of SptP–GFP into culture supernatants of *S. typhimurium*. SptP–GFP was not

secreted into the culture supernatant of *S. typhimurium*, indicating that the inability of InvC to unfold the SptP–GFP chimera renders this protein non-competent for secretion (Fig. 4c). In contrast, fusion of the targeting signals of SptP to PhoA, which is thermodynamically less stable²⁴, resulted in a chimera that was efficiently unfolded by InvC (Fig. 4d) and was efficiently secreted by the *S. typhimurium* TTS pathway (Fig. 4e). These results indicate that proteins that are destined to be secreted by the type III pathway must

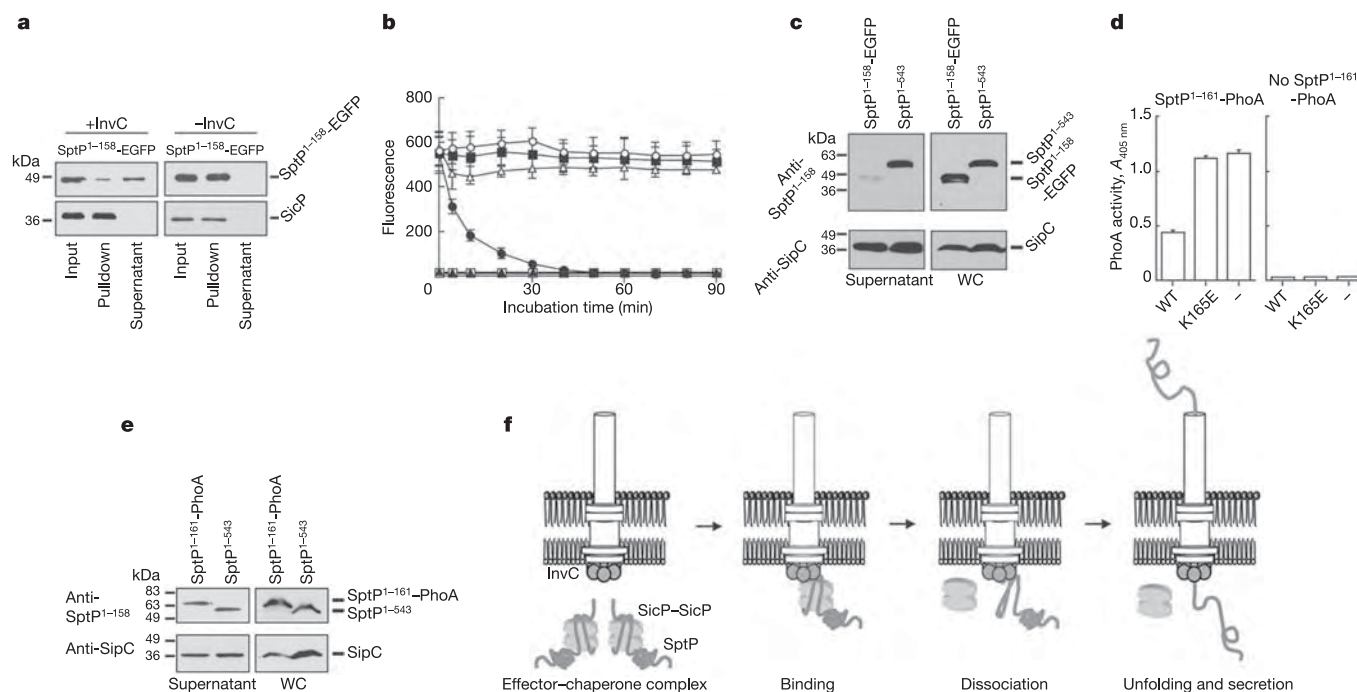


Figure 4 | Substrate unfolding is required for TTS. **a**, InvC was added to the SptP¹⁻¹⁵⁸–GFP protein complex in the presence of ATP; the release of SptP¹⁻¹⁵⁸–GFP was assayed as described in Fig. 2. **b**, InvC or its catalytic mutant InvC^{K165E} was added to purified SicP–SptP¹⁻¹⁵⁸–GFP complex and the GFP fluorescence (λ_{ex} = 488 nm, λ_{em} = 538 nm) of the samples was measured over time. Results are means ± s.d. for three independent measurements. Open circles, InvC^{K165E} plus SicP–SptP¹⁻¹⁵⁸–EGFP; filled squares, SicP–SptP¹⁻¹⁵⁸–EGFP; open triangles, InvC plus SicP–SptP¹⁻¹⁵⁸–EGFP; filled circles, SicP–SptP¹⁻¹⁵⁸–EGFP plus 6 M guanidinium chloride; open squares, InvC; filled triangles, InvC^{K165E}; open diamonds, buffer. **c**, Top, culture supernatants and whole-cell lysates

(WC) from *S. typhimurium* SB1404 expressing SicP along with wild-type SptP or SptP¹⁻¹⁵⁸–GFP were separated by SDS–PAGE and analysed for the presence of the different proteins by western immunoblot analysis. Bottom, as a control, blots were re-probed to examine the presence of the type III secreted protein SipC. **d**, InvC or its catalytic mutant InvC^{K165E} was added to the SicP–SptP¹⁻¹⁶¹–PhoA complex and the phosphatase activity was measured as indicated in Methods. Results are means ± s.d. for three independent measurements. **e**, The secretion of SptP¹⁻¹⁶¹–PhoA into culture supernatants of *S. typhimurium* SB1404 was examined as indicated above. **f**, Model for TTS (see the text for details).

be 'primed' for the ATPase-catalysed unfolding. Indeed, some features of the crystal structure of the enzymatic domains of SptP are consistent with this hypothesis. For example, the GAP domain region immediately adjacent to the chaperone-binding domain seemed highly disordered¹⁸. In addition, the better-ordered portion of the GAP domain stabilizes an extended-loop region at the N terminus of the PTP domain. It is therefore possible that, once the GAP domain has been unravelled, the loss of support for this extended-loop structure could facilitate the unravelling of the entire PTP domain.

Our results show a central role for the TTSS-associated ATPases in hitherto poorly understood steps of protein secretion, such as substrate recognition and chaperone release from, and unfolding of, type III secreted proteins (Fig. 4f). Our results also revealed remarkable parallels between InvC and the function of triple AAA+ ATPases, a family of proteins involved in a variety of biological processes^{25,26}. Although the TTSS-associated ATPases have not previously been considered to be members of this protein family²⁷, the ability of InvC to organize in hexameric rings, to disassemble protein complexes and to recognize and unfold protein substrates closely resembles equivalent activities of AAA+ ATPase disassembly machines^{25,28}. The similarities seem to extend to essential features of the substrate recognition process, because TTSSs and AAA+ machines recognize a poorly conserved set of unstructured short peptide signals located at the N terminus of the target proteins. TTS of virulence factors and the translocation of substrates into triple AAA+ machines might therefore have a common evolutionary origin and might consequently share mechanistic features.

METHODS

Bacterial strains. *Escherichia coli* XL-1 Blue (Invitrogen) was used for DNA manipulations and *E. coli* BL-21 (DE3) was used for the expression of recombinant proteins. The *Salmonella enterica* serovar Typhimurium SB1404, which carries a Δ sicP-sptP deletion, was constructed by allelic exchange as described previously²⁹.

Plasmids. Bicistronic plasmids encoding GST-SicP-SptP¹⁻¹⁵⁸, GST-SicP-SptP¹⁻⁵⁴³, GST-SicP-SptP¹⁻¹⁵⁸-EGFP and GST-SicP-SptP¹⁻¹⁶¹-PhoA³⁵⁻⁴⁷² were constructed by standard recombinant DNA techniques with the expression plasmid pGEX-KG as a backbone. Plasmids expressing GST-SptP¹⁶¹⁻⁵⁴³ (ref. 18), polyhistidine-tagged InvC and its mutant forms¹⁶, or polyhistidine-tagged CdtA, a subunit of the cytolethal distending toxin³⁰, have been described previously.

Purification of recombinant proteins. Recombinant polyhistidine-tagged monomeric and hexameric InvC were purified as described elsewhere¹⁶. Monomeric and hexameric forms of purified InvC were separated by gel-filtration chromatography as described previously¹⁶. GST-SicP in complex with SptP¹⁻¹⁵⁸, SptP¹⁻⁵⁴³ or the chimeric protein SptP¹⁻¹⁵⁸-EGFP or SptP¹⁻¹⁶¹-PhoA³⁵⁻⁴⁷² was purified from *E. coli* carrying the respective bicistronic expression plasmids as described previously⁷.

Affinity chromatography binding assay. Purified GST-SicP-SptP¹⁻¹⁵⁸ complex was incubated with purified polyhistidine-tagged wild-type InvC or its mutants (InvC^{V28M}, InvC^{K165E} or InvC^{L376P}) in a binding buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM MgCl₂, 0.05% Tween 20, 0.001% BSA) in the presence of 100 μ M ATP or ATP- γ S for 30 min at 4 °C. Glutathione-Sepharose 4 FastFlow beads were added to the mixtures and incubated for 1 h at 4 °C. After the beads had been washed to remove unbound proteins, the samples were subjected to SDS-PAGE. Proteins were detected by western blotting with monoclonal antibodies directed against the different relevant proteins.

Chaperone release assay. Purified SicP-GST-SptP¹⁻¹⁵⁸ complex was absorbed with glutathione-Sepharose 4B FastFlow at 37 °C for 1 h. After the beads had been washed to remove unbound complex, wild-type InvC or InvC^{K165E} was added to the beads in the releasing buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM MgCl₂, 0.05% Tween 20, 0.001% BSA) in the presence of 100 μ M ATP or ATP- γ S. After gentle rocking at 4 °C for 1 h, samples were separated into supernatant and beads by centrifugation. Supernatant was precipitated with trichloroacetic acid. The resulting samples were subjected to SDS-PAGE, and proteins were detected by western immunoblotting.

Protease susceptibility assay. Purified SicP-SptP¹⁻⁵⁴³ complex was incubated with wild-type InvC or InvC^{K165E} in reaction buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM MgCl₂, 0.05% Tween 20, 0.001% BSA, 100 μ M ATP or 100 μ M ATP- γ S, as indicated) at 37 °C. At several time points, different amounts

of subtilisin (10⁻⁵ to 10⁻⁸ U) were added to the samples and incubated on ice for 30 min. To stop the protease reaction, 1 mM PMSF was added to each tube. All the samples were subjected to SDS-PAGE and proteins were detected by western blotting.

Protein tyrosine and alkaline phosphatase assays. The purified SicP-SptP¹⁻⁵⁴³ complex or SptP¹⁶¹⁻⁵⁴³ was mixed with the purified hexameric form of wild-type InvC or the catalytic InvC^{K165E} mutant at a molar ratio of 1:10 and incubated at 37 °C for 30 min; the tyrosine phosphatase activity was measured with a RediPlate 96 EnzChek tyrosine phosphatase assay kit (Molecular Probes) in accordance with the manufacturer's recommendations. The purified SicP-SptP¹⁻¹⁵⁸-PhoA³⁵⁻⁴⁷² was mixed with the purified hexameric form of wild-type InvC or the catalytic InvC^{K165E} mutant as described above and the phosphatase activity was measured with the substrate *p*-nitrophenyl phosphate with the use of standard procedures.

EGFP fluorescence assay. Purified GST-SicP-SptP¹⁻¹⁵⁸-EGFP complex was incubated with wild-type InvC or the catalytic InvC^{K165E} mutant in assay buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM MgCl₂, 100 μ M ATP, 0.05% Tween 20, 0.001% BSA) at 37 °C for different durations and the fluorescence (excitation wavelength 488 nm, emission wavelength 538 nm) in the samples was measured in a fluorimeter (Molecular Devices).

Received 5 April; accepted 1 July 2005.

- Galán, J. E. & Collmer, A. Type III secretion machines: bacterial devices for protein delivery into host cells. *Science* **284**, 1322-1328 (1999).
- Eichelberg, K., Ginocchio, C. & Galán, J. E. Molecular and functional characterization of the *Salmonella typhimurium* invasion genes invB and invC: Homology of InvC to the F₀F₁ ATPase family of proteins. *J. Bacteriol.* **176**, 4501-4510 (1994).
- Cornelis, G. R. & Van Gijsegem, F. Assembly and function of type III secretory systems. *Annu. Rev. Microbiol.* **54**, 735-774 (2000).
- Kubori, T. *et al.* Supramolecular structure of the *Salmonella typhimurium* type III protein secretion system. *Science* **280**, 602-605 (1998).
- Marlovits, T. C. *et al.* Structural insights into the assembly of the type III secretion needle complex. *Science* **306**, 1040-1042 (2004).
- Stebbins, C. E. & Galán, J. E. Priming virulence factors for delivery into the host. *Nature Rev. Mol. Biol.* **4**, 738-743 (2003).
- Stebbins, C. E. & Galán, J. E. Maintenance of an unfolded polypeptide by a cognate chaperone in bacterial type III secretion. *Nature* **414**, 77-81 (2001).
- Birtalan, S. C., Phillips, R. M. & Ghosh, P. Three-dimensional secretion signals in chaperone-effector complexes of bacterial pathogens. *Mol. Cell* **9**, 971-980 (2002).
- Michiels, T. & Cornelis, G. R. Secretion of hybrid proteins by the *Yersinia* Yop export system. *J. Bacteriol.* **173**, 1677-1685 (1991).
- Ramamurthi, K. S. & Schneewind, O. Substrate recognition by the *Yersinia* type III protein secretion machinery. *Mol. Microbiol.* **50**, 1095-1102 (2003).
- Lloyd, S. A., Forsberg, A., Wolf-Watz, H. & Francis, M. S. Targeting exported substrates to the *Yersinia* TTSS: different functions for different signals? *Trends Microbiol.* **9**, 367-371 (2001).
- Galán, J. E. *Salmonella* interaction with host cells: Type III secretion at work. *Annu. Rev. Cell Dev. Biol.* **17**, 53-86 (2001).
- Dreyfus, G., Williams, A. W., Kawagishi, I. & Macnab, R. M. Genetic and biochemical analysis of *Salmonella typhimurium* Flil, a flagellar protein related to the catalytic subunit of the F₀F₁ ATPase and to virulence proteins of mammalian and plant pathogens. *J. Bacteriol.* **175**, 3131-3138 (1993).
- Pozidis, C. *et al.* Type III protein translocator: HrcN is a peripheral ATPase that is activated by oligomerization. *J. Biol. Chem.* **278**, 25816-25824 (2003).
- Claret, L., Calder, S. R., Higgins, M. & Hughes, C. Oligomerization and activation of the Flil ATPase central to bacterial flagellum assembly. *Mol. Microbiol.* **48**, 1349-1355 (2003).
- Akeda, Y. & Galán, J. E. Genetic analysis of the *Salmonella enterica* type III secretion-associated ATPase InvC defines discrete functional domains. *J. Bacteriol.* **186**, 2402-2412 (2004).
- Kaniga, K., Uralil, J., Bliska, J. B. & Galán, J. E. A secreted tyrosine phosphatase with modular effector domains encoded by the bacterial pathogen *Salmonella typhimurium*. *Mol. Microbiol.* **21**, 633-641 (1996).
- Stebbins, C. E. & Galán, J. E. Modulation of host signalling by a bacterial mimic: structure of the *Salmonella* effector SptP bound to Rac1. *Mol. Cell* **6**, 1449-1460 (2000).
- Gauthier, A. & Finlay, B. B. Translocated intimin receptor and its chaperone interact with ATPase of the type III secretion apparatus of enteropathogenic *Escherichia coli*. *J. Bacteriol.* **185**, 6747-6755 (2003).
- Luo, Y. *et al.* Structural and biochemical characterization of the type III secretion chaperones CesT and SigE. *Nature Struct. Biol.* **8**, 1031-1036 (2001).
- Farr, G., Scharl, E., Schumacher, R., Sondek, S. & Horwich, A. Chaperonin-mediated folding in the eukaryotic cytosol proceeds through rounds of release of native and nonnative forms. *Cell* **89**, 927-937 (1997).
- Lee, S. H. & Galán, J. E. *Salmonella* type III secretion-associated chaperones confer secretion-pathway specificity. *Mol. Microbiol.* **51**, 483-495 (2004).
- Lee, V. T. & Schneewind, O. Yop fusions to tightly folded protein domains and

- their effects on *Yersinia enterocolitica* type III secretion. *J. Bacteriol.* **184**, 3740–3745 (2002).
24. Fischer, C., Schauerte, J., Wisser, K., Steel, D. & Gafni, A. Differences in the pathways for unfolding and hydrogen exchange among mutants of *Escherichia coli* alkaline phosphatase. *Biochim. Biophys. Acta* **1545**, 96–103 (2001).
 25. Sauer, R. *et al.* Sculpting the proteome with AAA⁺ proteases and disassembly machines. *Cell* **119**, 9–18 (2004).
 26. Ogura, T. & Wilkinson, A. AAA + superfamily ATPases: common structure—diverse function. *Genes Cells* **6**, 575–597 (2001).
 27. Frickey, T. & Lupas, A. Phylogenetic analysis of AAA proteins. *J. Struct. Biol.* **146**, 2–10 (2004).
 28. Dougan, D., Mogk, A., Zeth, K., Turgay, K. & Bukau, B. AAA + proteins and substrate recognition, it all depends on their partner in crime. *FEBS Lett.* **529**, 6–10 (2002).
 29. Kaniga, K., Bossio, J. C. & Galán, J. E. The *Salmonella typhimurium* invasion genes *invF* and *invG* encode homologues to the PulD and AraC family of proteins. *Mol. Microbiol.* **13**, 555–568 (1994).
 30. Lara-Tejero, M. & Galán, J. E. CdtA, CdtB, and CdtC form a tripartite complex that is required for cytolethal distending toxin activity. *Infect. Immun.* **69**, 4358–4365 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Horwich and G. Farr for providing us with GroEL^{D87K}, S.-H. Lee for plasmid constructs, and members of the Galán laboratory for critical reading of the manuscript. This work was supported by a Public Health Service Grant.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.E.G. (jorge.galan@yale.edu).

LETTERS

Direct observation of steps in rotation of the bacterial flagellar motor

Yoshiyuki Sowa^{1*}, Alexander D. Rowe^{2*}, Mark C. Leake², Toshiharu Yakushi³, Michio Homma³, Akihiko Ishijima^{1,4} & Richard M. Berry²

The bacterial flagellar motor is a rotary molecular machine that rotates the helical filaments that propel many species of swimming bacteria^{1,2}. The rotor is a set of rings up to 45 nm in diameter in the cytoplasmic membrane³; the stator contains about ten torque-generating units anchored to the cell wall at the perimeter of the rotor^{4,5}. The free-energy source for the motor is an inward-directed electrochemical gradient of ions across the cytoplasmic membrane, the protonmotive force or sodium-motive force for H⁺-driven and Na⁺-driven motors, respectively. Here we demonstrate a stepping motion of a Na⁺-driven chimaeric flagellar motor in *Escherichia coli*⁶ at low sodium-motive force and with controlled expression of a small number of torque-generating units. We observe 26 steps per revolution, which is consistent with the periodicity of the ring of FlgG protein, the proposed site of torque generation on the rotor^{7,8}. Backwards steps despite the absence of the flagellar switching protein CheY indicate a small change in free energy per step, similar to that of a single ion transit.

Direct observation of steps in ATP-driven molecular motors has revealed much about the fundamental mechanisms of these protein machines^{9–11}. Steps characteristic of the ATP-driven F₁ part of ATP synthase have been observed when the whole enzyme is driven by ion flow in the F₀ part¹², but steps have never been seen in a purely ion-driven molecular machine. Steps have not previously been resolved in the flagellar motor¹³ because of its high speed^{14,15}, the presence of multiple stator units in a single motor⁴ and the small step size predicted—either from structural data on rotor periodicity or from mechanical data obtained by dividing the free energy of one ion crossing the membrane by the torque that the motor generates. We took several measures to overcome these factors. We detected rotation by back-focal-plane (BFP) interferometry of 500-nm diameter polystyrene beads attached to spontaneously sticky flagellar filaments of *E. coli*, as described previously¹⁶, or by high-speed video recording of 200-nm fluorescent beads attached in the same way (Fig. 1b, c, inset). The structure and function of H⁺ and Na⁺ motors are similar, and functional chimaeras have been made containing different mixtures of H⁺ and Na⁺ motor components¹⁷. We slowed motor rotation to the point at which steps could be resolved by decreasing the sodium-motive force (SMF) in *E. coli* cells containing Na⁺-driven chimaeric motors (in which the H⁺-driven stator proteins MotA and MotB are replaced by the Na⁺-driven PomA and the chimaeric fusion protein PotB, respectively⁶; Fig. 1a). We decreased the SMF by a low external Na⁺ concentration in experiments with BFP detection and by cumulative photodamage in fluorescence experiments. We controlled the number of stator units by inducible expression of stator proteins in a strain with

wild-type stator proteins deleted. We also deleted *cheY* and *pilA* to avoid possible complications due to motor switching or sticking of beads to cell bodies, respectively. We obtained rotation speed by means of power spectra of the combined *x* and *y* signals¹⁶, and bead angles by modelling the raw data as the projection of an oblique circular orbit¹⁸.

Fast Na⁺-dependent rotation and reduced speed when Li⁺ replaced Na⁺ (data not shown) indicate that the chimaera behaves as a typical Na⁺ motor^{15,19}. For step experiments, we grew cells with low inducer levels to produce fully energized motors with a small initial number of stator units (typically three or four; Supplementary Information). Reduction of the SMF, either by lowering the Na⁺ concentration in BFP experiments (Supplementary Information) or by photodamage in fluorescence experiments, led to a decrease in both the number of stator units and the speed per unit. Figure 1b shows the speed of a 500-nm bead in a typical BFP experiment. The Na⁺ concentration was reduced from 5 mM to 0.1 mM for the interval between the black arrows, during which the motor rotated at about 1 Hz. This cell rotated with four stator units when first observed in 5 mM Na⁺, but this number was decreased to one by a previous reduction of [Na⁺] (data not shown). Discrete speed increments identical to those occurring after the induced expression of stator proteins¹⁶ were typical during recovery after the transient

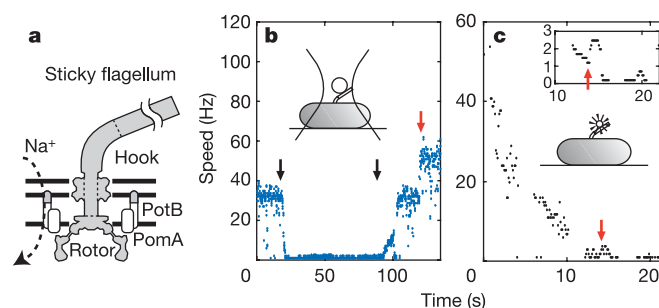


Figure 1 | Rotation measurements of chimaeric Na⁺-driven flagellar motors in *E. coli*. **a**, Diagram of the Na⁺-driven chimaeric flagellar motor. Components derived from *V. alginolyticus* (stators; PomA, PotB) and *E. coli* (PotB, rotor) are indicated in white and grey, respectively. **b**, Reducing SMF and motor speed by lowering the external Na⁺ concentration (from 5 mM to 0.1 mM and back; black arrows), with BFP detection (inset). **c**, Reducing SMF and motor speed by photodamage, with fluorescent detection (inset, lower). The speed doublings in **b** and **c** (red arrows; shown with expanded scales in **c**, inset, top) indicate a change from one to two stators. Data window lengths are 1 s (main figures) and 4 s (inset).

¹Department of Applied Physics, Graduate School of Engineering, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8603, Japan. ²The Clarendon Laboratory, Department of Physics, Oxford University, Parks Road, Oxford OX1 3PU, UK. ³Division of Biological Science, Graduate School of Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8602, Japan. ⁴PRESTO, Japan Science and Technology Corporation (JST) 4-1-8, Honmachi, Kawagoe, Saitama 332-0012, Japan.

*These authors contributed equally to this work.

removal of Na^+ (Fig. 1b), indicating reversible inactivation of stator units on disruption of the SMF. Similar phenomena have been reported for H^+ -driven motors of *E. coli*²⁰ and *Rhodobacter sphaeroides*²¹. Figure 1c shows the speed of a 200-nm fluorescent bead in a typical fluorescence experiment. We attribute periods of smooth speed reduction to photodamage-induced reduction of the membrane voltage²², and discrete speed changes (inset) to the consequent reversible inactivation of stator units. Speed doublings such as those in Fig. 1b, c were interpreted as a change from one to two stators. However, faster speed fluctuations were common and we were not always able to assign a definite stator number.

Figure 2a shows a sequence of images of a rotating 200-nm fluorescent bead with calculated bead centres superimposed (Methods). Figure 2b shows stepping rotation of 500-nm plain and 200-nm fluorescent beads attached to chimaeric flagellar motors. The diameter of bead orbits (insets) is consistent with attachment to a short flagellar filament stub. Steps were detected at speeds below 7 and 40 Hz in BFP and fluorescence experiments, respectively, and were not restricted to episodes with an estimated single stator unit. Figure 3a shows expanded sections of the traces in Fig. 2b (more examples are given in Supplementary Information), with the output of a step-finding algorithm (Methods) superimposed. Backwards steps were observed in both BFP and fluorescence experiments, with higher probability at lower speeds. As the strain lacks the switch-inducing protein CheY and never rotated backwards at high Na^+ concentration, backwards steps represent microscopic reversibility

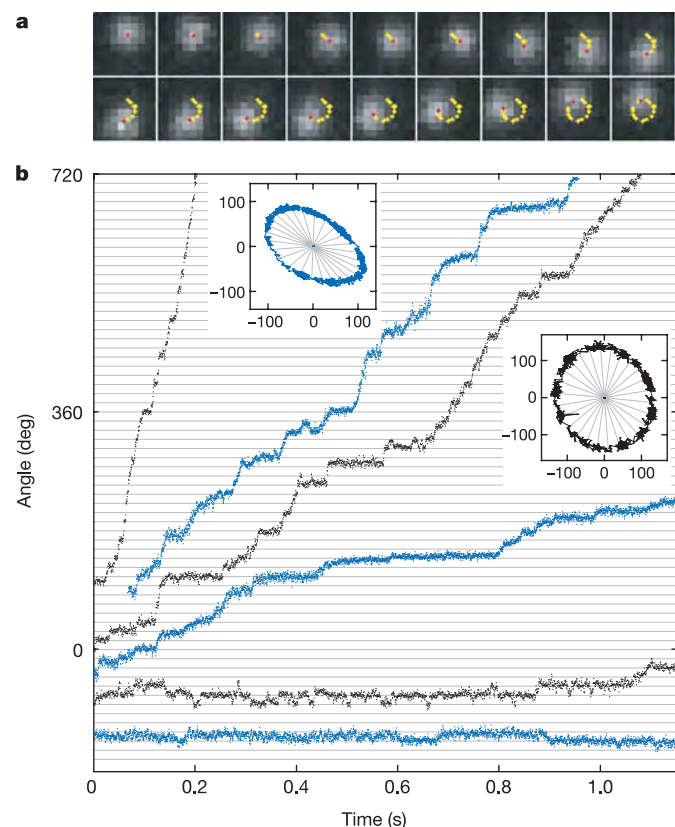


Figure 2 | Stepping rotation. **a**, Selected frames at 21-ms intervals during one rotation of a 200-nm fluorescent bead. Red and yellow dots mark the calculated bead centre in the current and previous images respectively. Each pixel is 80 nm square. **b**, Stepping rotation of flagellar motors with a range of average speeds. Bead positions are shown in the insets (scales in nanometres); bead angles are plotted against time in the main figure. Grey lines indicate $(360/26)^\circ$ in both main and inset figures. Blue and black traces are from BFP and fluorescence experiments, respectively. Backwards and forwards steps were observed at all speeds.

rather than motor switching. Similar steps have been seen, although less frequently, in ATP-driven molecular motors^{10,11}. Figure 3b shows several revolutions of a 200-nm bead, the histogram of all dwell angles during those revolutions and the power spectrum of that histogram. The peak at 26 per revolution in the spectrum corresponds to steps of 13.8° and indicates that successive revolutions show the same stopping angles.

We combined and analysed step data from different cells and both experimental techniques. Figure 4a shows the step-size histogram for all steps found. Assuming that larger steps are in fact unresolved multiple steps, we fit the distribution as a sum of gaussian distributions with means equal to integer multiples of step size, allowing different sizes and variances for forward and backward fits. The fitted step sizes are 13.7° (26 per revolution) and -10.3° (35 per revolution) for forward and backward steps, respectively. The difference between step sizes may be due to reorientation of the rotation axis on reversal, or it may be an artefact of measurement and analysis. Figure 4b shows the sum of histogram power spectra, similar to those of Fig. 3b, for all step data. Histograms of the levels between steps found by the step algorithm were also made; the sum of their power spectra is shown in the inset to Fig. 4b. The most striking peak is at 26 per revolution, with others at 11, 16 and 23 per revolution. After subdivision of the data, neither step size nor periodicity depended on the individual cell, experimental method, angle, estimated number of stator units or average speed of rotation.

Stepping motion in ATP-driven molecular motors reflects both the discrete molecular nature of the fuel and the periodicity of the 'track' along which the motor runs⁹⁻¹². Our observation of 26 steps per

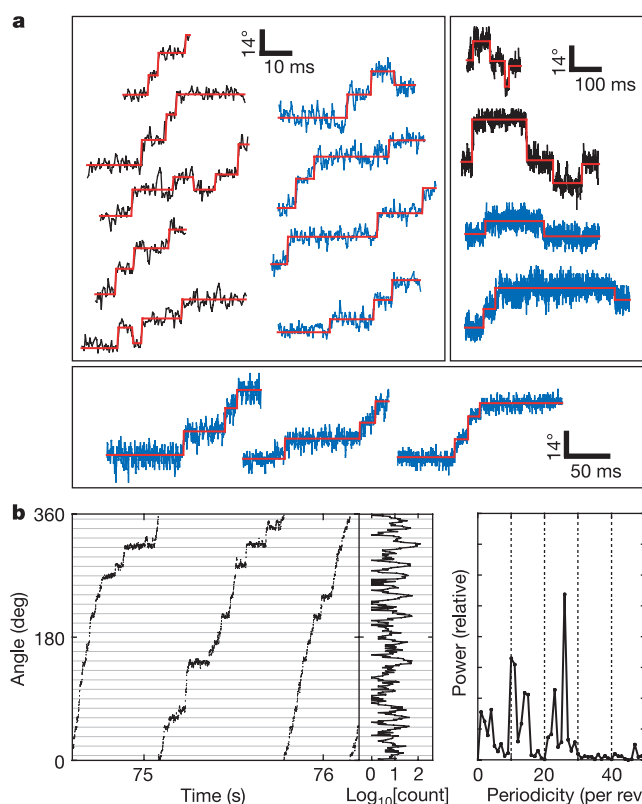


Figure 3 | Analysis of step size and periodicity. **a**, Selected sections of the traces in Fig. 2b, with the output of a step-finding algorithm superimposed (red). Blue and black traces are from BFP and fluorescence experiments, respectively, as in Fig. 2b. The dominant step size is about 14° . **b**, Plot of angle against time during three revolutions of a 200-nm bead, a histogram of dwell angles for the same revolutions and the power spectrum of that histogram. The peak at 26 per revolution corresponds to a step size of 13.8° , and shows that the motor stops at the same angles on successive revolutions.

revolution is consistent with the periodicity of the ring of Flg protein, the track on the rotor where torque is generated^{7,8}. The 10-fold or 11-fold periodicity matches the filament and hook²³; other periodicities observed may be due to interactions between other components of the flagellum with as-yet undetermined symmetry. Some cells in BFP experiments rotated at about 1 Hz even in 5 mM Na⁺. Steps were seen in these unusually slow episodes (data not shown), similar to or possibly slightly smaller than 1/26 of a revolution. Locking of one stator in one state of its mechanochemical cycle while other stators continued normally could produce steps with the periodicity of the locked state. Unusually slow episodes are excluded from our data analysis, but we cannot entirely rule out a possible contribution of a locking effect. Interval length distributions for data with a narrow range of average speed were sometimes, but not always, single-exponential. The apparent independence of step size on stator number, non-exponential interval length distributions and the observation of occasional steps smaller than 1/26 of a revolution may require the revision of existing models of the flagellar motor as a set of independent poisson-stepping stators^{16,24}.

Our experiments cannot determine whether single steps correspond to single ion transits. Regardless of the nature of the stepping process, the ratio of forwards to backwards step probabilities is $\exp(-\Delta G/kT)$, where ΔG is the free energy available to drive one step and kT is the average thermal energy. In our experiments ΔG ranged from 0 to $3kT$, depending on rotation speed. This is equivalent to the free energy of one ion transit through a SMF of up to -75 mV. Measuring the SMF in *E. coli* under these conditions would reveal whether the observed steps could be driven by single-ion transits. However, previous data indicate that single ions in fully energized wild-type motors cannot drive steps as large as 14° . If about ten torque-generating units pass 1,200 H⁺ ions per revolution²⁵, then one ion in one unit should step about 3° , assuming tightly coupled independent units¹⁶. Energy conservation sets an upper bound to the angle coupled to one ion transit, equal to (free energy per ion)/(maximum torque per unit). A wild-type *E. coli* cell with a

protonmotive force of about 150 mV ($6kT$ per ion)²⁶ drives a 1- μ m bead with an estimated 280 pN nm per unit¹⁶, giving an upper bound of about 5° . It is possible that stoichiometry changes at low SMF or stator number or that smaller substeps will be resolved in future. Alternatively, a mechanical step may be coupled to several ion transits, requiring the accumulation either of ions or of mechanical strain between steps. The latter mechanism is believed to occur in ATP-synthase, coupling three or four ion-driven steps in F_O to a single ATP-synthesizing step in F₁. By combining high-resolution measurements of flagellar rotation with manipulation of the SMF, the chimaeric flagellar motor described here will enable a detailed comparison to be made between the mechanism of the bacterial flagellar motor and those of its ATP-driven relatives.

METHODS

***E. coli* chimaera.** *E. coli* strain YS34 ($\Delta cheY$, $fliC::Tn10$, $\Delta pilA$, $\Delta motAmotB$) was derived from strain RP4979 (ref. 27) ($\Delta cheY$) and transformed with plasmids pYS11 (*fliC* sticky filaments²⁸, ampicillin resistance, pBR322 derivative) and pYS13 (*pomA* *potB* (ref. 6), isopropyl β -D-thiogalactoside inducible, chloramphenicol resistance, pMMB206 derivative). Deletions of *pilA* and *motAB* were made as described in ref. 29; *fliC::Tn10* was transduced from HCB1271 (ref. 16). Cells were grown for 5 h at 30°C from frozen stocks, with shaking, in tryptone broth (TB) containing antibiotics to preserve plasmids. Isopropyl β -D-thiogalactoside (1–10 μM) was added for low-level expression of stator proteins.

Rotation measurement. All experiments were performed at 23°C . Speeds were obtained from power spectra of combined (x,y) data¹⁶, using data windows of length 1 or 4 s beginning at intervals of 0.1 s. Motor angles were obtained by fitting an ellipse to bead trajectories under the assumption that trajectories represent the projection of a circular orbit onto the focal plane of the microscope¹⁸.

For BFP experiments, polystyrene beads (diameter 535 nm; Polysciences) were attached to truncated flagella of immobilized cells, and bead position (x,y) was measured by BFP laser interferometry (sample rate 4 kHz) as described¹⁶. NaCl (or LiCl) was added to the Na⁺-free motility medium in various concentrations; the total concentration of added salt was adjusted to 85 mM with KCl. Custom-made flow chambers allowed a complete exchange of medium in about 5 s.

For fluorescence experiments, fluorescent polystyrene beads (diameter 227 nm, 'yellow/green'; Molecular Probes) were attached as above. A mercury-arc lamp provided epifluorescence excitation (475-nm bandpass) at an intensity of about 1 W cm^{-2} . Images (16×16 pixels, each 80 nm square in the sample plane; 505 nm long-pass) were sampled at a frame rate of 2.4 kHz with a cooled, back-thinned, electron-multiplying charge-coupled device camera (iXon DV860-BI, Andor Technology). Bead position was determined with a precision of about 5 nm by a two-dimensional gaussian fit to the bead image.

Step resolution. The relaxation time of a 500-nm bead attached to an elastic hook rotating about an axis 150 nm from its diameter is about 1.1 ms (ref. 30). For a 200-nm bead the time is about 0.15 ms. These times are lower limits assuming no contribution from the flagellar stub¹⁶, and they set upper limits of about 900 and 6,500 to the number of steps that can be detected per second in BFP and fluorescence experiments respectively. Our actual limits were lower, namely about 200 s^{-1} and about $1,000\text{ s}^{-1}$, respectively. The latter limit was attained by analysis of dwell-time histograms.

Step analysis. A computer algorithm divided angle versus time records into straight-line portions corresponding to episodes of constant average speed. A second algorithm based on the method of J. W. Kerssemakers (personal communication) detected steps as follows. First, an entire episode was divided into two intervals at the point giving the best least-squares fit to a single-step function. Second, an excess of steps (N_{max} per revolution on average) was assigned by repeatedly dividing, as in the first step, the interval containing the largest range of angles in the previous iteration. Third, a 'quality factor' Q was defined for each assigned step as $Q^2 = (x_1 - x_2)^2 / [(var_1/n_1) + (var_2/n_2)]$, where x_i , var_i and n_i are the mean angle, the variance and the number of points in an interval, respectively, and $i = 1$ and $i = 2$ indicate intervals immediately before and after a step. The lowest-quality step was removed and adjacent intervals were concatenated, but only if the quality was below a threshold Q_{min} . Last, the previous step was repeated until no steps remained with $Q < Q_{\text{min}}$. The accuracy and sensitivity of the algorithm were tested and a suitable value of Q_{min} for each episode was chosen by applying the algorithm to simulated data with $N_0 = 10, 20, 30, 40$ and 80 Poisson-distributed steps per revolution and brownian noise and average speed similar to real data.

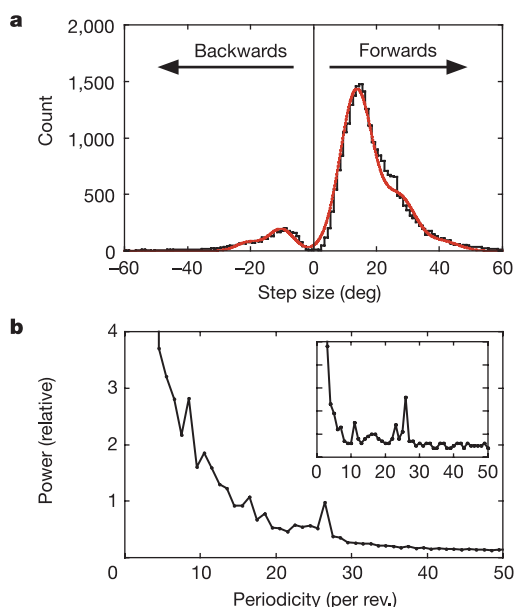


Figure 4 | Summary of step analysis. **a**, A histogram of step sizes found in all episodes of approximately constant speed (1,400 revolutions, 9 different cells, 28,611 steps). A multiple-gaussian fit (red line) gives step sizes of 13.7° ($\sigma = 5.16^\circ$) and -10.9° ($\sigma = 3.9^\circ$) for forward and backward steps, respectively. **b**, The sum of histogram power spectra (see Fig. 3b) for the same data set. Inset, summed spectra of histograms of mean levels between found steps. The forward step size is in close agreement with the peak in the summed spectra, confirming that there are 26 steps per revolution.

Received 1 April; accepted 11 July 2005.

1. Berry, R. M. & Armitage, J. P. The bacterial flagella motor. *Adv. Microb. Physiol.* **41**, 291–337 (1999).
2. Berg, H. C. The rotary motor of bacterial flagella. *Annu. Rev. Biochem.* **72**, 19–54 (2003).
3. Thomas, D. R., Morgan, D. G. & DeRosier, D. J. Rotational symmetry of the C ring and a mechanism for the flagellar rotary motor. *Proc. Natl Acad. Sci. USA* **96**, 10134–10139 (1999).
4. Blair, D. F. & Berg, H. C. Restoration of torque in defective flagellar motors. *Science* **242**, 1678–1681 (1988).
5. Berry, R. M., Turner, L. & Berg, H. C. Mechanical limits of bacterial flagellar motors probed by electrorotation. *Biophys. J.* **69**, 280–286 (1995).
6. Asai, Y., Yakushi, T., Kawagishi, I. & Homma, M. Ion-coupling determinants of Na⁺-driven and H⁺-driven flagellar motors. *J. Mol. Biol.* **327**, 453–463 (2003).
7. Suzuki, H., Yonekura, K. & Namba, K. Structure of the rotor of the bacterial flagellar motor revealed by electron cryomicroscopy and single-particle image analysis. *J. Mol. Biol.* **337**, 105–113 (2004).
8. Lloyd, S. A. & Blair, D. F. Charged residues of the rotor protein Flg essential for torque generation in the flagellar motor of *Escherichia coli*. *J. Mol. Biol.* **266**, 733–744 (1997).
9. Mehta, A. D. *et al.* Myosin-V is a processive actin-based motor. *Nature* **400**, 590–593 (1999).
10. Schnitzer, M. J. & Block, S. M. Kinesin hydrolyses one ATP per 8-nm step. *Nature* **388**, 386–390 (1997).
11. Yasuda, R., Noji, H., Kinosita, K. Jr & Yoshida, M. F₁-ATPase is a highly efficient molecular motor that rotates with discrete 120 degree steps. *Cell* **93**, 1117–1124 (1998).
12. Diez, M. *et al.* Proton-powered subunit rotation in single membrane-bound F₀F₁-ATP synthase. *Nature Struct. Mol. Biol.* **11**, 135–141 (2004).
13. Berg, H. C. in *Cell Motility* Vol. A (eds Goldman, R., Pollard, T. & Rosenbaum, J.) 47–56 (Cold Spring Harbor Press, New York, 1976).
14. Berg, H. C. & Turner, L. Torque generated by the flagellar motor of *Escherichia coli*. *Biophys. J.* **65**, 2201–2216 (1993).
15. Sowa, Y., Hotta, H., Homma, M. & Ishijima, A. Torque-speed relationship of the Na⁺-driven flagellar motor of *Vibrio alginolyticus*. *J. Mol. Biol.* **327**, 1043–1051 (2003).
16. Ryu, W. S., Berry, R. M. & Berg, H. C. Torque-generating units of the flagellar motor of *Escherichia coli* have a high duty ratio. *Nature* **403**, 444–447 (2000).
17. Yorimitsu, T. & Homma, M. Na⁺-driven flagellar motor of *Vibrio*. *Biochim. Biophys. Acta* **1505**, 82–93 (2001).
18. Yasuda, R., Noji, H., Yoshida, M., Kinosita, K. Jr & Itoh, H. Resolution of distinct rotational substeps by submillisecond kinetic analysis of F₁-ATPase. *Nature* **410**, 898–904 (2001).
19. Liu, J. Z., Dapice, M. & Khan, S. Ion selectivity of the *Vibrio alginolyticus* flagellar motor. *J. Bacteriol.* **172**, 5236–5244 (1990).
20. Fung, D. C. & Berg, H. C. Powering the flagellar motor of *Escherichia coli* with an external voltage source. *Nature* **375**, 809–812 (1995).
21. Armitage, J. P. & Evans, M. C. Control of the protonmotive force in *Rhodospseudomonas sphaeroides* in the light and dark and its effect on the initiation of flagellar rotation. *Biochim. Biophys. Acta* **806**, 42–55 (1985).
22. Neuman, K. C., Chadd, E. H., Liou, G. F., Bergman, K. & Block, S. M. Characterization of photodamage to *Escherichia coli* in optical traps. *Biophys. J.* **77**, 2856–2863 (1999).
23. Samatey, F. A. *et al.* Structure of the bacterial flagellar hook and implication for the molecular universal joint mechanism. *Nature* **431**, 1062–1068 (2004).
24. Samuel, A. D. & Berg, H. C. Torque-generating units of the bacterial flagellar motor step independently. *Biophys. J.* **71**, 918–923 (1996).
25. Meister, M., Lowe, G. & Berg, H. C. The proton flux through the bacterial flagellar motor. *Cell* **49**, 643–650 (1987).
26. Gabel, C. V. & Berg, H. C. The speed of the flagellar rotary motor of *Escherichia coli* varies linearly with protonmotive force. *Proc. Natl Acad. Sci. USA* **100**, 8748–8751 (2003).
27. Scharf, B. E., Fahrner, K. A., Turner, L. & Berg, H. C. Control of direction of flagellar rotation in bacterial chemotaxis. *Proc. Natl Acad. Sci. USA* **95**, 201–206 (1998).
28. Kuwajima, G. Construction of a minimum-size functional flagellin of *Escherichia coli*. *J. Bacteriol.* **170**, 3305–3309 (1988).
29. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA* **97**, 6640–6645 (2000).
30. Block, S. M., Blair, D. F. & Berg, H. C. Compliance of bacterial flagella measured with optical tweezers. *Nature* **338**, 514–518 (1989).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank H. Berg and K. Fahrner for the gift of strain HCB1271. The research of R.B., M.L. and A.R. was supported by combined UK research councils through an Interdisciplinary Research Collaboration in Bionanotechnology, that of A.I., M.H. and T.Y. by Grants-in-Aid from the Ministry of Education, Science, Sports, Culture and Technology of Japan, that of M.H. and T.Y. by Soft Nano-Machine Project of JST, and that of Y.S. by JSPS Research Fellowships for Young Scientists.

Author Contributions BFP experiments were performed by Y.S. and A.R., fluorescence experiments by A.R. and M.L., experimental design was by R.B., A.I., A.R. and Y.S., data analysis by R.B., Y.S. and A.R., and strain construction by Y.S., T.Y. and M.H. Y.S. and A.R. contributed equally to this work.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.M.B. (r.berry1@physics.ox.ac.uk).

ERRATUM

doi:10.1038/nature04218

Deficiency of *glutaredoxin 5* reveals Fe-S clusters are required for vertebrate haem synthesis

Rebecca A. Wingert, Jenna L. Galloway, Bruce Barut, Helen Foott, Paula Fraenkel, Jennifer L. Axe, Gerhard J. Weber, Kimberly Dooley, Alan J. Davidson, Bettina Schmid, Barry H. Paw, George C. Shaw, Paul Kingsley, James Palis, Heidi Schubert, Opal Chen, Jerry Kaplan, The Tübingen 2000 Screen Consortium & Leonard I. Zon

Nature 436, 1035–1039 (2005)

In this Letter, the surname of Bettina Schmid was incorrectly spelled as Schmidt. In addition, her affiliation should have been the Adolf-Butenandt-Institute, Department of Biochemistry, Schillerstrasse 44, D-80336 Munich, Germany, not her previous affiliation (the Stem Cell Program and Division Hematology/Oncology Children's Hospital and Dana-Farber Cancer Institute) as published.

-  FOCUS
-  SPOTLIGHT
-  RECRUITMENT
-  ANNOUNCEMENTS
-  EVENTS

naturejobs

Out in the cold

If the representation of women in science is bad, then that for ethnic minorities in research can only be called abysmal. Women, long outnumbered by men in physics and engineering, have at least been making some inroads in the life sciences. And although the 'glass ceiling' that keeps many from top positions still exists, some cracks are beginning to form. In the United States, progress for African-Americans and Hispanics is much less clear-cut, despite some programmes meant to address the problem.

Figures from the US National Science Foundation bear this out. African-Americans, Hispanics and other ethnic groups make up 24% of the US population, but they hold only 7% of scientific jobs. The numbers are even more skewed at the PhD level. In 1999, 381,600 white PhD scientists held jobs in the United States compared with 11,600 African-Americans and 12,900 Hispanics.

This summer, a grant from the National Institutes of Health was renewed to address this imbalance. The programme provides \$6.1 million to the Federation of American Societies for Experimental Biology to support a variety of student, postdoctoral and faculty

development initiatives, including mentoring projects and travel grants.

Although this grant is better than nothing, it seems a pittance compared with the size of the problem. Fortunately, some grass-roots organizations have sprung up to fill the gap. One of these is the African-American Male Achievers Network at the University of California, Los Angeles. This programme sends scientists to primary-school and high-school classrooms, then brings minority students into university laboratories (see *Nature* **434**, 418; 2005).

Other universities might want to consider instituting similar programmes, and send scientists of various ethnicities and nationalities into schools. Such outreach should be a part of every scientist's job description.



Paul Smaglik, *Naturejobs* editor

CONTACTS

Publisher: Ben Crowe
Editor: Paul Smaglik
Marketing Manager: David Bowen

European Head Office, London
 The Macmillan Building, 4 Crinan Street
 London N1 9XW, UK
 Tel: +44 (0) 20 7843 4961
 Fax: +44 (0) 20 7843 4996
 e-mail: naturejobs@nature.com

Naturejobs Sales Director:
 Nevin Bayoumi (4978)
European Sales Manager:
 Andy Douglas (4975)

Natureevents: Sille Opstrup (4994)
UK/RoW/Ireland/Italy:
 Nils Moeller (4953)
 Irene Viglia-Atton (4944)
Scandinavia/Spain/Portugal:
 Evelina Rubio Håkansson (4973)
France/Switzerland/Belgium:
 Amelie Pequignot (4974)
Germany/Austria/The Netherlands:
 Reya Silao (4970)

Advertising Production Manager:
 Billie Franklin
 To send materials use London
 address above.
 Tel: +44 (0) 20 7843 4814

Fax: +44 (0) 20 7843 4996
 e-mail: naturejobs@nature.com
Naturejobs web development: Tom Hancock
Naturejobs online production: Niamh Shields

European Satellite Office
Germany/Austria/
The Netherlands:
 Patrick Phelan
 Tel: +49 89 54 90 57 11
 Fax: +49 89 54 90 57 20
 e-mail: p.phelan@nature.com

US Head Office, New York
 345 Park Avenue South,
 10th Floor, New York,

NY 10010-1707
 Tel: +1 800 989 7718
 Fax: +1 800 989 7103
 e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

Japan Head Office, Tokyo
 Chiyoda Building,
 2-37 Ichigayatamachi,
 Shinjuku-ku,
 Tokyo 162-0843
 Tel: +81 3 3267 8751
 Fax: +81 3 3267 8746
Asia-Pacific Sales Director: Rinoko Asami
 e-mail: rasami@naturejpn.com

To do today

IMAGES.COM/CORBIS

Moving research quickly forward to publication tops the pre-tenure 'to do' list. **Kendall Powell** ticks off project management tips.

Who hasn't written a task on their to-do list just as it is completed, so they can still experience the joy of crossing it off? Kim Barrett, a physiologist at the University of California, San Diego, finds it no less satisfying to check things off the to-do list on the digital organizer that goes with her everywhere.

"I've always been a list maker — not just of the big things, but of the intermediate steps too, outlining what needs to get done to get a publication out," says Barrett, who is vice-chair for research at the university's department of medicine. Writing lists is the main way investigators keep track of their research projects and goals.

But new investigators, hit simultaneously with piled-on administrative responsibilities and freedom to pursue their research interests, may find it increasingly difficult to track all lab activities, either mentally or with a clutch of sticky notes. Successful project management involves directing many research threads without letting individual experiments, ideas or trainees slip through the cracks.

"You want your team focused and moving ahead, producing publishable results as quickly and as frequently as possible," says Federico Rosei, a nanomaterials scientist at the University of Quebec, Canada. He and other project-juggling experts share their methods of tracking projects, organizing lab staff and expediting the final product: they use paper, programs and people to achieve that highly desired 'p' — publication.

Paper

Structured project management is needed when you have "project strings going in several directions at once," says Victoria McGovern, programme officer for infectious diseases at the Burroughs Wellcome Fund in Research Triangle Park, North Carolina. "It ensures that important things don't get shoved to the back of the bench when more urgent things come up," she says.

Most investigators keep a running list, in mind or on paper, of tasks to move research forward. They use e-mails, voicemails and calendar alarm programs to remind them of deadlines and administrative duties. Rosei admits that he often pulls off the highway to scribble down ideas on the back of any available

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

LAB RULES TO LIVE BY

Gillian Wu, an immunologist and dean at York University in Toronto, runs a research group of about 15. To keep projects going, she demands high-quality data and no-nonsense workers.

She sets these tough standards for herself:

“Always make a list of what is most important for today and do it. Don't let anything get in your way.”

“I make graduate students do the first draft of a paper and tell them I want to see it in a week — no extensions.”

“Make sure you only give projects the time that's needed and not a minute more.”

...and for her lab members:

“Remind me of this later — if you don't get back to me, I won't get back to you.”

“No sloppy experiments or trial runs are allowed. Which gel do you want to publish?”

“You should be able to turn to page 59 in your lab notebook and tell me what you did that day.”

“We're in this together. I'll help you get your PhD but you have to do your part.” K.P.



K. WALSH

writing surface. Continually updated paper lists can eventually become an outline for a manuscript.

But McGovern warns that keeping lists without a specific time course in mind can lead to thinking: "I'm going to get all these things done because I've written them down." Instead, she says, "Write them down with a calendar in front of you." Otherwise you are likely to get frustrated when teaching or family activities interrupt your science, or vice versa.

Shri Kulkarni, an astronomer at the California Institute of Technology in Pasadena, keeps an hour-by-hour web-based calendar roughly scheduled a year ahead to track important conferences and grant deadlines and let colleagues know "which part of the world I am in and where to call me".

Barrett and McGovern advise new investigators to take some quiet time to plan for project advancement, using the first three months of a year to do a detailed organization of the lab's agenda for the next nine months, or half an hour each morning (before trainees arrive) to plan what has to be done that day.

Programs

As the size of a lab and the number of administrative duties increase, keeping all those details in mind becomes a burden. For the overtaxed brain or for very complex research projects, McGovern recommends formal project-management software.

These programs allow you to break a project into smaller tasks and figure out whose expertise and what materials each piece requires. "It lets you tie together pieces of work that make sense and the dependencies between them," McGovern says.

The programs automatically redraw flow charts and reset deadlines when something unexpected comes up. McGovern was "geekishly excited" to find out the latest version of her software communicated with her e-mail program better, so "things don't fall off my radar screen". One of the most useful aspects, she says, is that the program helps you identify which parts of a project can move forward even if another part is stalled.

People

Another way to free up brain capacity is to delegate responsibility for remembering the details to senior students and postdocs. Recruiting the best people is arguably the first step in getting projects rolling fast.

"The main difference between projects that go well and those that stall are the people," says Arshad Desai, a cell biologist in the school of medicine at the University of California, San Diego. "Young principal investigators feel they should recruit anyone, but it's wise to try to recruit the right people for projects."

Kulkarni puts third-year graduate students in charge of every aspect of writing and submitting proposals and manuscripts. His lab staff also have carte blanche to order items costing less than \$1,000. "It's not just a time-saving device on my part," he explains. "It's a way to win their respect." It's important for lab members to learn that research, like any other job, requires that certain tasks get finished by the end of each day.

Kulkarni admits that his philosophy does not result in immediate improvements in efficiency, but says younger lab heads should be "willing to play longer bets". Those bets produce fourth- and fifth-year students who are valuable assets in meeting lab goals.

When Gillian Wu became dean of science and



Help fortune to smile on you: don't wait until you think the story is complete before you start writing, say Federico Rosei (top) and Kim Barrett.

engineering at York University in Toronto, she had to get tough on efficiency (see 'Lab rules to live by', opposite). She has strict rules limiting Internet surfing and political discussions, and her 'right-hand' lab manager keeps on top of day-to-day progress. Because Wu spends a lot of time away from the lab, weekly meetings are mandatory for all.

"They have to bring their own primary data, not just words or histograms," says Wu. A principal investigator at any stage has to look at the primary data to catch unanticipated results that might be important for project direction, she adds.

Desai holds group meetings lasting 60–90 minutes, and confers with the presenting student individually for an hour or two afterwards. "That gives me a three-to-four-hour chunk to just think about that specific project," he says, and students get his undivided attention on a regular basis. Many investigators keep folders for each person's project to review and update at lab meetings. But people invariably get stuck, technically or conceptually, and a good adviser must help them navigate through.

Wu and Barrett stress that a new lab head should stay at the bench for as long as possible to understand the daily rhythm and technical requirements of projects. "You actually have more quality control of the data and a sense of how to move things along," says Barrett. "There will be time enough to sit at the computer." She also tries to foster an environment in which staff can freely express the hurdles holding them up. Hearing the 'bad news' early, she notes, might be critical to understanding the scientific process under study.

Wu demands clean, clear notebooks so problems can be solved by going back through data line by line. She also expects all experiments to be carried out to publishable quality. Both those rules smooth the transition to writing a paper — with ready-made materials and methods, results and figures.

Publication

Knowing when a project has reached the point when a paper can be crafted is tricky, experts admit. Rosei says an introduction can be written in advance, as you should know why you are pursuing the project before starting work. He urges researchers not to delay writing until they have a 'complete' story.

"It is impossible to put the word 'end' to anything, and you don't want to be scooped because someone else had the guts to go out with the incomplete story," he warns. As soon as there is a coherent story, he says, have the student write a first draft.

Barrett agrees, and begins a paper "once I've got a sense that there's some flesh on the bones of the hypothesis". She instructs trainees not to delay to do 'loose-end' experiments that can be added during review. And she combats inertia by giving students a deadline for a small, easy section, such as methods, first.

Desai notes that writing papers with students, especially focusing on the discussion section, gives him a good opportunity to review the literature. He gets his students to keep up with the finer literature details.

"Not having too many details in your head helps you see the big picture and new insights," Desai says. After all, at the top of every researchers' career-long to-do list is guiding ideas into reality.

Kendall Powell is a freelance science writer based in Broomfield, Colorado.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

IMAGES.COM/CORBIS

MOVERS

Geoffrey West, president, Santa Fe Institute, Santa Fe, New Mexico



2003–present: Distinguished professor, Santa Fe Institute, Santa Fe, New Mexico

1999–present: Professor of biology, University of New Mexico, Albuquerque, New Mexico

1982–present: High-energy physicist, Los Alamos National Laboratory, Los Alamos, New Mexico

Geoffrey West is fluent in the language of the Universe. He has harnessed the power of mathematics to tackle some of biology's most fundamental questions, which all too often go unnoticed by an academic community moving towards applied research in order to secure funding.

Realizing, as a young boy sitting on the white cliffs of his native England, that he could easily determine the distance to the horizon using passing ships shaped his future study of physics. After graduating from the University of Cambridge, he found inspiration in a PhD assistantship at Stanford University in California. At Stanford, it was the collegial interactions with fellow classmates, rather than his instructors, that motivated him to continue in science. His time in California also paved the way for this theoretical physicist to cross disciplinary boundaries.

In his spare time, West began to contemplate the biological process of ageing and to work on a way of quantifying what keeps organisms alive. His growing interest in biology led him to the friendly, open atmosphere of New Mexico's Los Alamos National Laboratory, which had a history of interdisciplinary work. In 1992, the proposed Superconducting Super Collider project he'd been involved with for years was scrapped, prompting him to do more than just dabble in biology. And an opportune encounter with a biologist helped him make that transition.

Mike Simmons, then vice-president of the Santa Fe Institute (SFI), a private, independent institute in New Mexico, introduced West to Jim Brown, an ecologist at the University of New Mexico who was looking to collaborate with a physicist. "I was speaking Latin and he was speaking Greek," West says of their meetings. It took a year of work to get past the jargon, reach across their scientific disciplines and publish a seminal paper in *Nature* detailing some of the laws that dictate an organism's metabolic rate.

"People don't realize you need to create an atmosphere in which researchers are willing to ask elementary questions without feeling defensive or vulnerable," West says.

Last year, West moved to the SFI. As its newly elected president, he will now be able to foster the environment he hoped to find as a college student — one that emphasizes big-picture science. "One of the great ironies is that despite talk of interdisciplinary research, it remains to be realized at the grassroots level," says West.

He hopes the SFI will continue to fill that void, providing a haven for researchers to explore without boundaries. ■
Virginia Gewin

RECRUITERS & ACADEMIA

Getting out of a rut

In my travels to academic institutions and companies around the world, I frequently meet scientists at varying stages of their careers who feel trapped and uninterested in their jobs. Some in their mid-50s feel they can't escape the 'publish-or-perish' rat race, and younger ones see no flexibility in the career path they have chosen.

The good news is that they can build tremendous flexibility and excitement into their careers by taking on a wider range of professional activities, such as editorial work, consulting and administration. Measured participation in such activities, even just on an extracurricular level, will not get in the way of doing good research, and can open up alternative career paths.

Thanks to this multitasking strategy, I have been recruited to a part-time position as chief scientific officer of iCo Therapeutics, a biotechnology company based in Vancouver, Canada. I have also been considered for jobs as editor-in-chief of journals and dean of arts and sciences at a major university. Many of my peers have had similar opportunities.

Indeed, many scientists have successfully made the transition into an allied field and are very happy as a consequence. Some have turned part-time work as industrial consultants or members of scientific advisory

boards into senior-level positions at biotechnology companies. Postdoctoral fellows can join these companies and work their way up.

Work as a reviewer can lead to memberships on editorial boards, and even to full-time appointments as editor-in-chief of major journals. Younger scientists can work at law firms, often receiving a fully financed legal education while training to be a patent lawyer.

Major service on university committees can lead to jobs as high-level university administrators, or at a more junior level to work at consultancies.

And scientists at all levels often become primary- or secondary-school teachers. Some join fast-track teaching programmes to qualify for posts at state schools, and others work at leading independent or private schools. This transition often begins by volunteering to lecture at such schools.

For every disenchanted scientist I have met in academia or industry, there are many others who have shown that, as long as scientists develop some skills in allied fields, there is truly no need to feel stuck in a rut. ■

Santa Jeremy Ono is a professor of biomedical science and associate dean at University College London.

GRADUATE JOURNAL

A time to reconnect

Suddenly, everything is over. A year of intensive study, the final examination, followed by celebrations, leafing through a year's worth of papers, cramming my life into eight boxes, and then going home to Germany for a month of relaxation before I begin my PhD.

It was a tough year, not just for me, but also for my family and, most of all, my girlfriend who lives in Germany. She and I tried to simulate a normal life as we managed to see each other at least every second weekend. Between these visits were many long-distance phone calls and long working hours on both sides. Nevertheless, our relationship remains intact and strong, and I'm very grateful for her support. These holidays are my girlfriend's and family's as much as they are mine.

Being at home has made me realize how important it is for graduate students to maintain a personal life outside the laboratory and to take time off to be with loved ones. We cannot forget that the people who love us guarantee us a home when we return from the lab, be it for a weekend or when our degree work is finally finished. And they are as affected by our career decisions as we are ourselves. Time is precious, and spending it with my family has been just as valuable to me as spending it at the lab bench. ■

Tobias Langenhan is a graduate student in neuroscience at the University of Oxford, UK.

Sandcastles: a dystopia

Caught up in a storm.

Kathryn Cramer

You had big summer plans to revive the mathematics career that might have been but wasn't. But there were your children, their eyes trained upon you. *Here we are. Entertain us.* You shoot endless camcorder footage of your son catching frogs; of your daughter taking off her shoes while you say in the background: "Don't you dare take off those shoes." At the end of the summer, you vacation with your mother-in-law, age 93. She watches TV while you, your husband and kids go to the beach.

You return to the cottage wanting to talk about children and their sandcastles and the rising tide; instead your mother-in-law wants to talk about the impending disaster. Unfortunately, your mother-in-law knows her hurricanes. And so you return home. You read the weather on the web in disbelief. Surely, this is like the comet Kohoutek: the real event will be the advance coverage? Nothing can live up to that; and of course, it doesn't. You dream that your great-grandmother's bones are floating out to sea. You aren't relieved, as you were in denial in the first place, but you begin to comprehend; and so you see it when it begins to happen.

You have watched children with their sandcastles and you can't get that image out of our mind. Water gradually overwhelming the defences, the castles collapse, surrender to the ocean. A levee breaks. You strive for understanding, find what you think are before-and-after pictures on disparate websites. You bring them together and tell your friends. You were wrong initially, but your attempts attract a swarm of helpful techies. Cameramen in 'copters snap photos that appear on the web, seemingly of 'the' break. Satellite images are at your fingertips. You compare, contrast, argue. You and your newfound friends, most of whom you've never met and whom you know only by handle, realize with growing horror that the situation is worse, much worse, than anyone has yet described. Not one break, but three or four. Thousands of people are going to die, but you don't yet dare say what you understand, for how could a housewife with a fancy computer set-up understand it before those who should be saving those people?

You begin to get specific addresses in the mail. *No one is helping, maybe you can? hello i was trying to find the condition of my home my address was...my*

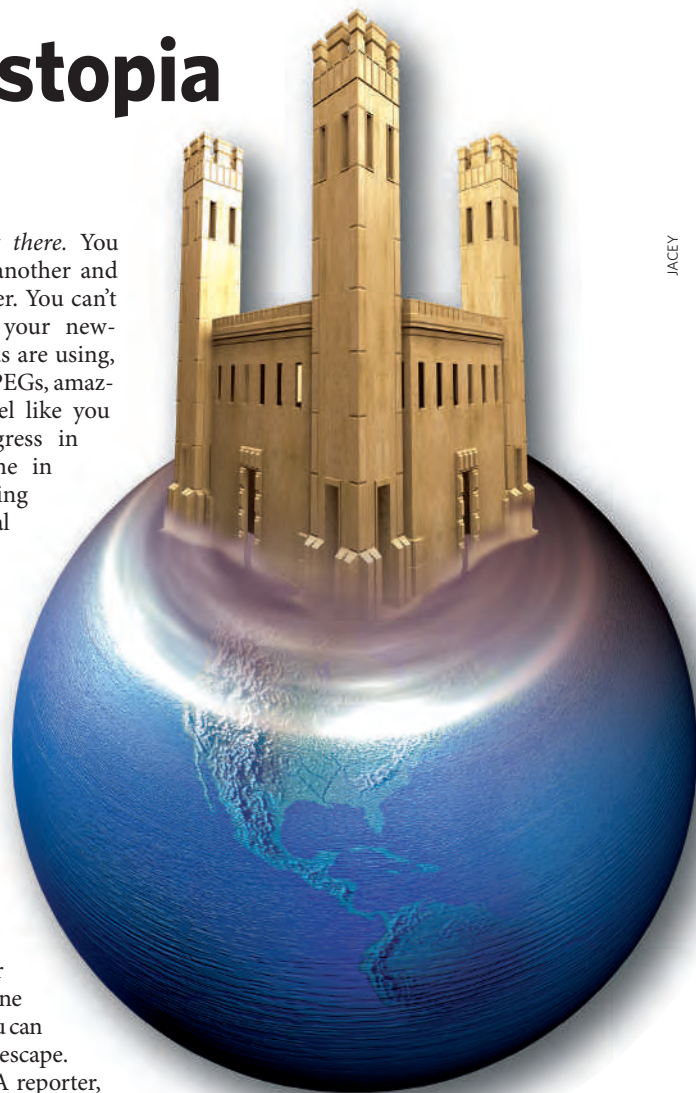
two dogs were left there. You answer that. And another and another and another. You can't run the software your new-found techie friends are using, but they send you JPEGs, amazing JPEGs. You feel like you are watching progress in Civil War medicine in fast-forward; amazing innovations in digital cartography are happening overnight, there for you when you wake up in the morning. And what you see then is beautiful and terrifying. You can't play too. All you can do is publish the images and make suggestions. And answer and answer and answer. Someone else asks you how you can help her relatives escape. You try your best. A reporter, impressed with the pictures, gets in touch, and you are interviewed.

When you arrive at your friends' house for the weekend, in a place so remote that cell phones don't work, you check your e-mail and discover many more requests for help, along with four requests for interviews. By the end of the weekend, you are world famous. The world is looking for good news, and for a few days, you are it — a bright spot in the bleakness. Meanwhile, the scenario you understood unfolds and unfolds and unfolds. You really did understand things long before those who are supposed to be keeping watch. It seems impossible, but you did. Many more people are dying and you know it, and you are doing what you can, but it is not enough, cannot be enough, because you are a smart housewife with a computer and all you can do is write.

And you write and write, because people want to know if their houses are under water, if their dogs or relatives or friends are dead. And they know you can help, because it said so in *The New York Times*. You have left politics out of it, because you want your help to be available to anyone, no matter what their beliefs. But at a certain

point that is no longer possible. What you begin to understand is that when you wrote that the dogs probably didn't drown because there was only 18 inches of water max. at that address, that you were terribly naive. The dogs will be shot; the house razed. But you don't say so, because having exchanged e-mails with so many, you think you know how they feel.

And the tide begins to turn: The Collective You is winning. The press won't be silenced after all; the remaining residents can stay; the *tabula rasa* scenario begins to recede. And you write and write. You have learned to respond, learned a kind of one-on-one compassion and respect that might otherwise have passed you by. But you cannot get back to your own world: you are living in the dystopia in which all hell breaks loose. And now you know. And you watch your children start the school year and you know where they live. ■ Kathryn Cramer is widely credited for having popularized the use of Google Earth and Google Maps for examining the damage inflicted by Hurricane Katrina.



JACEY